January 2020

Molecular Mechanisms Of Transcriptional Responses In Stress-Induced Human Cells

Sayantani Ghosh Dastidar

MOLECULAR MECHANISMS OF TRANSCRIPTIONAL RESPONSES IN STRESS-INDUCED HUMAN CELLS

by

Sayantani Ghosh Dastidar
Bachelor of Science, Zoology, University of Calcutta, 2010
Master of Science, Zoology, University of Calcutta, 2012

A Dissertation
Submitted to the Graduate Faculty

of the

University of North Dakota

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota

May
2020

This dissertation, submitted by Sayantani Ghosh Dastidar in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Sergei Nechaev, Chairperson

Min Wu, Co-Chairperson

Colin Combs, PhD

L. Keith Henry, PhD

Manu Manu, PhD

This dissertation is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

Chris Nelson, PhD
Assoc. Dean of the School of Graduate Studies

3/13/20
Date

PERMISSION

Title: Molecular Mechanisms of Transcriptional Responses in Stress-Induced Human Cells

Department: Biomedical Sciences

Degree: Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work, or in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Sayantani Ghosh Dastidar
May, 2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

"*It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness…*" – Charles Dickens, A Tale of Two Cities.

This journey of half a decade has been transformational for me. These years have helped me not only to grow as a scientist but also to grow as an individual. I would like to thank each and everyone who had, in their own ways, contributed to my growth.

I would like to thank my advisor Dr. Sergei Nechaev for believing in me. He has always pushed me out of my comfort zone and instilled in me fearlessness to tread into the unexplored. He has given me the freedom to independently address fundamental questions of transcriptional biology. He has always encouraged me to practice leadership roles in lab and outside, have worked with me tirelessly during every phase of grad school, and have nurtured in me the ability to think critically. His excellent problem-solving skills and pursuit of cutting-edge techniques are qualities which I have always adored. This journey would have been impossible without his constant support and guidance.

I would like to thank my co-advisor, Dr. Min Wu, for his excellent guidance. He has been very generous with his time and always met with me whenever I needed him. His advice had often helped me handle difficult situations that I had faced during my PhD.

I thank my committee members – Dr. Colin Combs, Dr. Keith Henry, and Dr. Manu Manu for being extremely generous with their time, for always attending all my talks and seminars, and their critical feedback on my projects.

I would like to thank all the professors in our Department of Biomedical Sciences for always giving me advice and encouragement. I would especially like to thank Dr. Barry Milavetz for the uncountable conversations in the corridor on how to think about cellular mechanisms, for generating interest in viruses, and for being so caring towards me. I want to thank Dr. Alexei Tulin for always making me feel as a part of his lab, allowing me to participate in his lab meetings as well as his lab socials, and for letting me work with his microscopes. I want to thank Dr. Motoki Takaku for his help during standardization of ATAC-sequencing experiments, and for the numerous conversations ranging from transcription factors to post-graduate travel plans, Dr. Junguk Hur for always helping me learn Bioinformatics and data analysis, Dr. Archana Dhasarathy for asking interesting questions in lab meetings, Dr. Thad Rosenberger for keeping a tab on my credits, and Dr. Othman Ghribi for his constant mental support.

I want to thank the Bioinformatics Core - Dr. Bony De Kumar, Hannah Ness, and Danielle Perley for their immense help, especially during the last two years, from sequencing my libraries to helping me learn genomic data analyses.

I want to thank my lab members, past and present, without whom I probably wouldn't have survived graduate school. They are family to me. I want to thank Dr. Nii Koney Kwaku Koney, Dr. Ann Samarakkody, Oscar Nnoli, Dr. Damien Parrello, Dr. Ata Abbas, Bo Lauckner, Maria Vlasenok, Maria Privratsky and all undergraduate students. I want to thank all graduate students in UND especially Smruthi Rudraraju, Dr. Jessica Warns, Gbolahan Bamgbose and my classmates of 2014. I want to thank Nafisa Ferdous for being my best friend ever and constantly supporting me through the journey. I want to thank all the post-doctoral fellows in the department, especially Dr. Atrayee Ray, Dr.

Iaroslava Karpova, Dr. Guillaume Bordet, and Dr. Harpreet Kaur for helping me in numerous aspects such as immunostaining, brainstorming ideas, organizing thesis, and above all for their kindness.

I want to thank all the Biomedical Sciences Staff especially Bonnie Kee for managing my grant and reimbursements and helping me immensely in formatting my thesis.

I want to thank all my family members and my friends for being very supportive during this time. I want to thank my childhood friends, Arundhati Sarkar, Sumana Maity and Supriya Sadhukhan for their unwavering faith in me. I want to thank my best friend Barnil Bhattacharjee for making me dream big and for being my happy place. I want to thank my friend Dr. Niladri Sinha for being a constant critic of my work and for his numerous one-liners such as “Laser focus: always signal over noise”.

I want to thank mum-mum, Mitali Dasgupta, and didi, Debasmita Nandi, for their encouragement and appreciation.

I want to thank my dadu, thakuma, and didan for being so proud of their only grand-daughter in whatever she did.

Finally, I would like to thank my Ma and Baba for always allowing me to be a free spirit, for all the sacrifices they have made for me, and for always respecting my choices. My father has always been practical in life and has instilled determination in me. My mother has dedicated all her life in bringing me up and nurturing me; she has instilled in me the virtue of imagination and discipline which have been invaluable assets for me during my PhD.

And, I want to thank my permanent ‘boyfriend’, Dr. Suman Dasgupta, who has patiently listened to each and every story of failure (which is probably 95% of the time), gave impromptu motivational speeches, ordered surprise pizzas during dark days, and actually asked important questions from a very different perspective that helped me approach the problems from a completely different angle.

This journey would not have been possible without the inspiration and motivation from all of you.

মা, বাবা ও সুমন কে

ABSTRACT

Precise regulation of gene expression is essential for maintaining cell homeostasis and survival. Rapid responses to intracellular or extracellular stimuli involve highly regulated genome-wide changes in the transcriptional landscape. However, works on transcriptional regulation thus far has been primarily focused on genes and gene promoters. Consequently, ubiquitous transcription outside of genes has been greatly neglected. Therefore, in order to understand genome-wide transcriptional changes with stimuli, we looked at transcription factor distribution, RNA Polymerase II (Pol II) dynamics, histone changes, chromatin accessibility at both promoters and promoter-distal regions along with changes in gene transcription. With this approach, we were able to create a more comprehensive picture of how cells respond when exposed to stimuli.

Previous studies in mouse and human cells have shown with heat stress at elevated temperature, there is massive genome-wide binding of transcription factor HSF1 (Heat-Shock Factor 1) (Mahat *et al.*, 2016; Vihervaara *et al.*, 2017). Genome-wide analyses revealed there are approximately 280,000 HSF1 binding elements in the human genome. However, with heat-shock, HSF1 binds to less than 4% of these sites indicating HSF1 binding might be context-dependent. To address whether HSF1 binding depends upon stimulus and/or cell-type, we exposed two different cell types, MCF7 breast cancer and K562 chronic myelogenous leukemia cells to a heat-dependent (heat stress at 43°C) and a heat-independent stress (Arsenic) and compared genome-wide changes in HSF1 distribution, histone marks, RNA Pol II and nascent transcription. We found HSF1 binds

across the entire human genome during HSR and is a highly sensitive indicator of global heat-shock response (HSR) among the marks we tested. This genome-wide HSF1 binding is independent of stress type but is dependent on cell type. This cell type specific difference in HSF1 binding is more prominent at distal intergenic regions rather than at promoters. The mechanism of this difference in HSF1 binding in different cell types might not be dependent on chromatin accessibility as detected by ATAC-sequencing but might be potentially pre-determined at ground state of cells maintaining cells in a poised state for rapid orchestration of HSR.

We next studied the dynamics of RNA Polymerase (Pol II) regulation and gene expression under different stresses and different human cell types. HSR has been associated with massive repression of gene expression due to elevated temperature (Mahat *et al.*, 2016; Vihervaara *et al.*, 2017). We found this repression is cell type dependent and takes place by distinct cell-specific mechanisms. Regulation of gene expression at the level of Pol II can occur at two steps: (i) during recruitment of Pol II at promoter and (ii) during release of Pol II from promoter-paused site to gene body. We found this mechanism of recruitment versus release at promoter differs between cell types and can, in turn, explain the cell-type specific extent of repression.

These studies indicated that distal intergenic regions or potential enhancer regions might be important regulatory elements showing significant transcriptional changes with stress. Therefore, we next utilized another biological model of IL-1β -induced inflammation in lung adenocarcinoma cells (A549) to investigate the role of enhancers in gene regulation. We found that rapid genome-wide transcriptional changes take place in A549 cells induced with IL-1β. Transcription at genes and their nearest predicted

enhancers display similar patterns of expression changes with inflammation. We further found that RNA production at genes and enhancers are not separated by time but are co-transcriptional events.

In conclusion, this dissertation demonstrates that cells undergo rapid and robust genome-wide transcriptional changes with inflammatory or environmental stresses affecting gene-promoters, gene-bodies as well as extragenic regulatory elements such as enhancers. The mechanisms of these transcriptional changes to stress responses in different cell types under different conditions might be determined during cell lineage specification and is probably dependent on the availability of cell-specific transcriptional regulators involved in the recruitment and release of Pol II at gene regulatory elements.

CHAPTER 1

INTRODUCTION

1.1 Motivation

Our bodies are exposed to various kinds of intracellular and extracellular or environmental stresses throughout our lifetime. As a response to these stresses, cells have evolved mechanisms to induce rapid changes in their molecular machinery to maintain homeostasis and fight these stresses. These stress-responses are extremely precise and well-regulated and, take place within seconds of exposure to the stress-inducing stimuli. Cells activate stress-specific pathways leading to synthesis of stress-response proteins that buffers the effect of stress and protect the cells from dying. Stress response mechanisms are present in almost all eukaryotic organisms studied, from yeast to mammals and play critical roles during development and cell-fate determination as well as in diseased states such as cancer. My dissertation focuses on investigating the mechanisms of rapid changes that cells undergo in their transcriptional landscape in response to the environmental and inflammatory stresses. In this chapter, I will start by discussing the fundamental concepts and later introduce the specific questions that I have addressed in my thesis.

1.2 Transcription

Transcription is the process wherein an RNA molecule is synthesized from the DNA template. Transcription is a highly regulated process involving enzymes and molecular machinery that is highly conserved from bacteria and archaea to eukaryotes[1].

The transcription cycle in all organisms involves three major steps, initiation, elongation and termination, after which the transcription machinery releases the DNA strand and might be recycled for the next round of transcription. Each of these steps are precisely regulated by multiple factors including nucleosome positioning and chromatin accessibility, gene regulatory elements such as promoters and enhancers, binding of transcription factors (TFs), and assemblage of the essential proteins to form the transcriptional machinery (Fig 1.1).

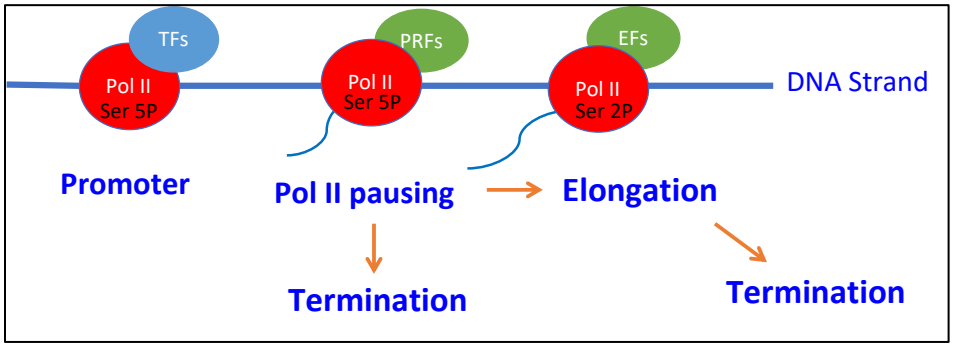


Figure 1. 1. Schematic representation of the process of transcription by RNA Polymerase II.

I would briefly elucidate each step of the transcription process and the regulatory check points in them, the key factors involved in transcription and the role of the chief gene-regulatory elements such as promoters and enhancers in transcription.

1.3 Regulatory Steps During Transcription

1.3.1 Transcription initiation

Initiation of transcription occurs from a defined region at the 5' end of the target gene also known as 'transcription start site' (TSS) that is present within the core promoter (explained later) where the transcriptional machinery binds and forms the transcription pre-

initiation complex (PIC). This PIC comprises RNA Polymerase II (Pol II) and general transcription factors (GTFs) such as TATA-binding protein (TBP), TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH[2]. The transcriptional machinery assembles on the promoter region, loosen and unwind DNA, place the template DNA strand in the active site of Pol II and the machinery start moving along the DNA producing an RNA transcript (Fig 1.2).

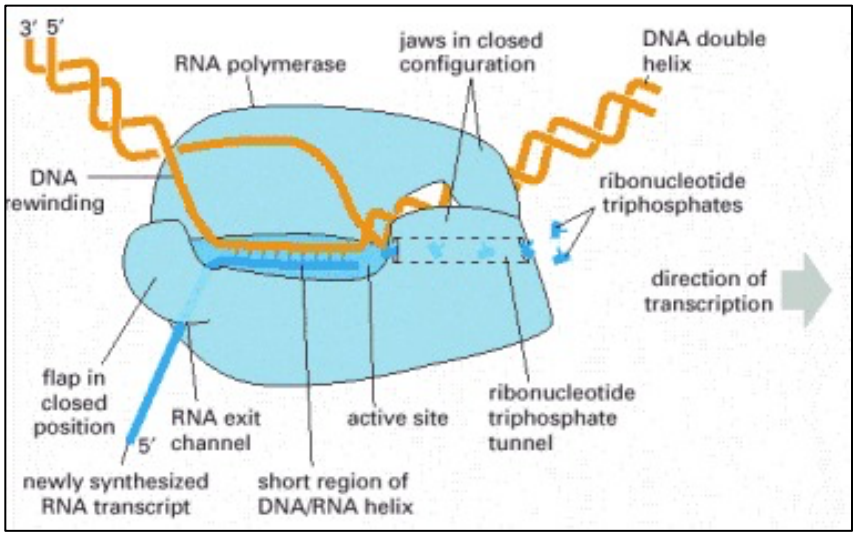


Figure 1. 2. Figure showing RNA Polymerase II in the act of transcription. *Figure taken from Molecular Biology of the Cell. 4th edition., Alberts B, Johnson A, Lewis J, et al., New York: Garland Science; 2002*

Recent studies show that transcription initiation is not a continuous process but takes place as discrete events called 'bursts'. 'Transcriptional bursting'[3,4,5,6] is a process in which genes either exist in an "ON" (burst initiation) state or an "OFF" (burst termination) state and RNAs are produced in discrete episodic bursts marked by long periods of inactivity[5]. Bursting can occur by various mechanisms. Single-molecule studies in *E. coli* suggest DNA topology is an important contributor for bursting phenomenon. According to Chong et al, 2014 bursting happens because of the positive and negative supercoils present within the DNA introducing torque during transcriptional elongation[7]. Asynchronous

generation of these coils and limitation of enzyme gyrase that relieves positive supercoiling by inserting DNA double-stranded breaks cause the transcriptional machinery to halt. Increase in gyrase concentration releases positive super-coiling resulting in transcriptional bursting. In eukaryotes, bursting is gene-specific and more complicated than prokayotes[8]. Mutations in TATA-box[9], in promoter regions[10], and in factors of pre-initiation complex can change the frequency and size of bursting[8]. A single event of bursting can produce from 2 RNA molecules to hundreds of RNA molecules[8]. Presence of transcription factors can increase bursting frequency by increasing recruitment of Pol II at TSS[11]. Transcription factors might also increase bursting frequency by facilitating enhancer-promoter interaction. This might explain how transcription factor binding leads to increased gene expression. However, the rate, frequency and size of bursting depends on multiple other factors such as cell size, chromatin complexity, promoter strength etc. and thus needs further investigation.

After initiation, transcription machinery encounters a second checkpoint known as Pol II pausing.

1.3.2 RNA Polymerase II (Pol II) Pausing

Pol II pausing is a key rate-limiting step in the process of transcription and a major checkpoint before elongation.

Evidence of Pol II pausing was first reported in early 1980's in human HeLa cells treated with 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazde (DRB)[12] and, in beta-globin gene of hen erythrocytes[13] where they observed premature termination of transcription and shorter RNA transcript[12,13] than normal-length mRNA. In 1986, David Gilmour and John Lis reported high concentrations of Pol II near the 5' end of heat-shock

gene hsp70. They saw Pol II accumulation mainly between -12 and +65 nucleotides relative to the TSS in inactive Schneider line 2 cells of *Drosophila melanogaster*[14]. However, with heat-shock induction, they noticed high levels of Pol II all across the gene[14] indicating that this Pol II stalling is temporary and is released when the gene becomes active. In 1990, Lis's laboratory showed that Pol II pausing is a widespread event occurring at other heat-shock responsive genes as well as other constitutively expressed genes[15]. In initial genome-wide studies, approximately 30% of genes were reported to display Pol II pausing and this fraction has been confirmed by performing nascent RNA analysis such as GRO-seq in various cells such as human lung fibroblasts [16,17], mouse embryonic stem (ES) cells [16,18] and *Drosophila* S2 cells [16,19]. However, depending on which method is used and in which system, Pol II has been found to be stalled on 30% -70% of actively transcribed genes at ground state[20,21,22]. Recently, it has been shown in rat neurons that Pol II pausing is almost always associated with all transcriptional events taking place at or outside of gene promoters[23].

Paused Pol II has been shown to be relatively stable compared to other steps of transcription, with a half-life of approximately 6 min in Drosophila or mammalian cells[24]. This stability allows for signal co-ordination from other transcription regulatory components. Based on the signals received, Pol II either continues to make a new transcript or aborts transcription. However, the mechanisms of how Pol II pausing is regulated – factors involved in pausing and pause-release, and consequences of Pol II pausing are still open questions in the field. Current understanding in the field is that Pol II pausing is mediated by *Ne*gative *E*longation *F*actor (NELF)[25] and *D*RB-*s*ensitivity *i*nducing *f*actor

(DSIF) acting together while pause release is mediated by *p*ositive *t*ranscription *e*longation *f*actor b (p-TEFb)[26], a component of super-elongation complex (SEC)[27].

NELF complex is a highly conserved transcriptional complex and are found to be present in most higher eukaryotes studied except *Caenorhabditis elegans, Saccharomyces cerevisiae,* and *Arabidopsis thaliana*[28,25]. NELF contains 4 sub-units – NELF-A, NELF-B, NELF-C/D, and NELF-E. The current model of Pol II pausing proposes that NELF along with other factors such as DSIF binds to Pol II and 'stalls' the elongating Pol II. P-TEFb, which comprises of Cdk9 kinase and one of the several cyclins (Cyclin T1, Cyclin T2a, Cyclin 2b, Cyclin K)[26] in mammals, phosphorylates NELF and Pol II CTD at Ser 2 causing it to be released from stalled Pol II which then resumes transcription[25,26]. However, there have been other factors that are recently being identified such as PAF1[29], additional members of Super elongation complex (SEC)[27] which are shown to either stabilize pausing or aid in pause release. The interplay of multiple regulatory factors makes Pol II pausing a highly complicated event more than just an ON/OFF switch needing to be carefully dissected and studied.

1.3.3 Transcription Elongation

Ironically, rather than being a repressive event, Pol II pausing is required for efficient transcriptional elongation in-vivo[30]. In Drosophila, partial knock-down of NELF with siRNA leads to decreased Pol II occupancy at promoters of active genes, and multiple promoters eventually get occupied by histones leading to their permanent repression[31]. Transcriptional elongation is the major step where a full-length RNA transcript is produced by Pol II and is also a highly regulated step of transcription. Single-molecule studies show that this rate of production of RNA varies between genes and can be almost threefold

different affecting mRNA accumulation. This variation in transcription rate can serve as a key regulatory step during various developmental processes[20]. In addition to differences among individual genes, the speed of Pol II elongation is not constant across the entire gene; near promoters Pol II transcribes at a rate of 0.5kb/minute whereas after approximately 15kb, the rate of elongation becomes 2-5kb/minute[20,32,33]. Pol II elongation is further regulated by co-transcriptional events including splicing of introns, mRNA cleavage and polyadenylation sites[20]. The time frame of Pol II elongation is consistent with the time taken for co-transcriptional splicing which is estimated to be around 20-30 seconds[20,21,34]. Pol II elongation is also additionally controlled by helicases such as RECQL5 which is known to directly interact with Pol II and stall its progression[35] whereas TFIIS and elongin help in Pol II progression[36].

1.3.4 Transcription Termination

Termination occurs when the nascent RNA transcript and RNA Pol II is released from DNA and is essential for proper formation of mature RNA products. Failure to terminate transcription can lead to serious consequences. For genes arranged in clusters or in divergent orientation, Pol II if not stopped, can slide from the downstream region of a target gene to the promoter region of the next gene interfering with its activity – a process called 'transcriptional interference'[37,38] ; the transcriptional machinery can also collide with DNA replication fork and lead to genomic instability[39]. For genes arranged in convergent orientation, failure to terminate may result in production of overlapping transcripts that can act as siRNA bringing down gene expression[40,41,42].

Transcription termination is an extremely complicated and well-regulated process. There are a few models on transcription termination. The 'torpedo' model of termination

states that Pol II continues to synthesize nascent RNA transcript even after cleavage at poly-adenylation site (PAS) at the 3'-end[39]. Nuclear XRN2, a 5' to 3' exonuclease, comes as a "torpedo" and degrades the upstream transcript in kinetic competition with the elongating Pol II finally catching up with the "battleship" which here is Pol II and dissociating it from DNA[39]. The 'allosteric' model of termination proposes elongating Pol II can sense when it is near an active PAS leading to conformational changes in the active site of the polymerase causing dissociation of the molecule from DNA[39]. At certain DNA sequences Pol II can, instead of moving forward, move backward. This event is called backtracking and has recently been found to be a widespread occurrence[43]. Backtracking can lead to disassembly of Pol II from DNA by inducing conformation changes of Pol II ultimately resulting in transcription termination[44].

1.4 Key Players of Transcription

1.4.1 RNA Polymerase II (Pol II)

Although the basic mechanisms of transcription remain the same, regulation of eukaryotic transcription is much more complex than in prokaryotes[1]. While bacteria and archaea have only one polymerase, eukaryotes have three polymerases (RNA Pol I, RNA Pol II, and RNA Pol III) to synthesize specific classes of RNA. However, most of the subunits of different polymerases are homologous to each other, implying that they have the same basic structure and function.

Pol II is the primary enzyme that is responsible for eukaryotic mRNA transcription. Transcription by RNA polymerase II is more complicated than transcription by any other polymerases involving nearly 60 polypeptides whereas the other polymerases requires only a few of these subunits. Different stages of transcription is precisely controlled by different

states of Pol II phosphorylation and dephosphorylation at the C-terminal domain (CTD) of its largest subunit RPB1 (Fig 1.3). The CTD of mammalian Pol II is composed of 52 repeats of the consensus sequence Y_1 S_2 P_3 T_4 S_5 P_6 S_7 and each hydroxyl-containing residue is phosphorylated or dephosphorylated at specific steps of transcription[45]. During the formation of transcription pre-initiation complex (PIC), Pol II is completely dephosphorylated[46,47]. However, during initiation and productive elongation, Pol II exhibits different phosphorylation states. The states of Pol II phosphorylation and the mediating enzymes are listed in the table below (Table 1.1).

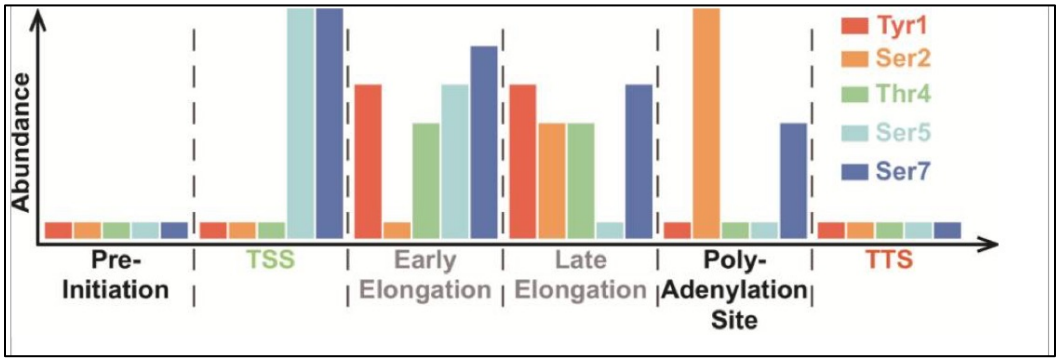


Figure 1. 3. ChIP-enrichment data of CTD phosphorylation residues in different stages of transcription in S. cerevisiae. The y-axis indicates approximate ChIP-enrichments found in genomic tiling arrays and the x-axis indicates approximate stage of transcription. Figure reused from *Mayfield et al., 2016 Biochim Biophys Acta. 2016 Apr: 1864(4): 382–387* with permission.

Table 1.1 Table illustrating the different states of Pol II during transcription and the mediating enzymes

Transcription Steps	Pol II CTD phosphorylation status	Mediating enzyme
Pre-initiation complex	Completely dephosphorylated/ No phosphorylation	CTD phosphatases
Initiation	Serine 5 (Ser 5), Serine 7 (Ser 7)	Cyclin-dependent kinase 7 (Cdk 7), a subunit of TFIIH
Pausing	Serine 5	Cdk 7
Productive elongation	Serine 2 (Ser 2)	Cdk 9 subunit of P-TEFb
Termination	Serine 2	Cdk 9
Backtracking	Serine 2	Cdk 9

1.4.2 Transcription Factors

Transcription factors (TFs) are an essential component of gene regulation. There are almost 1500 to 2000 TFs that are encoded by human genome[48,49]. TFs can be broadly divided into two groups: the first group control transcription during (i) initiation and the second group that control during (ii) elongation[50]. However, some TFs can control transcription during both initiation and elongation. TFs bind to their cognate DNA-binding motifs, attract co-factors that otherwise cannot bind to DNA, and mediate activation or repression of transcription[50]. TFs can bind to gene-regulatory elements such as promoters and enhancers as well as other regulatory-regions of DNA such as silencers and insulators and control gene expression[51]. Generally, eukaryotic TF recognize small 6-12bp long degenerate DNA sequences[51] suggesting that there are other modules of regulation, more than just sequence-specificity that govern TF binding to a site. It has been shown that the time of TF occupancy at a particular locus and the region that it binds to can play significant role in controlling gene expression[51]. For example, in *Drosophila melanogaster* mesoderm-specific TFs are present all along during development. But they have been shown to bind

only particular sets of enhancers at specific developmental stages, implying the temporal specificity with which TFs bind genomic regions[52].

Multiple TFs can bind to regulatory regions using different mechanisms that can result in very different outputs.

(i) Co-operative binding of TF: When TFs bind to adjacent sites on DNA, often facilitated by protein-protein interaction between them. This can produce a binary on-off switch-like effect and this pattern of TF binding is prevalent during determination of cell lineage. For example, during early Drosophila development, Bicoid TF expression is sharply controlled and binds only above a certain threshold leading to precise expression of only Bicoid target genes[51,53].

(ii) Assisted loading or collaborative competition model of TF binding: Competition between TFs for the same site might result in net increase of TF binding by inhibiting nucleosome repositioning[54].

(iii) Changes in chromatin conformation: TFs can lead to local DNA bending bringing about conformational changes of chromatin causing distant regulatory elements like enhancers and promoters to come closer to each other [51,55,56].

(iv) Pioneer factor TFs: These TFs are a special class of recently identified TFs that has been shown to possess some unique properties. They can interact with their binding motifs present on nucleosome-occupied chromatin and can induce changes in chromatin accessibility. They melt chromatin regions and can increase local chromatin openness helping other reprogramming factors to bind to the open region. They act as major deterministic factors in cell-fate decisions[48,57,58]. For example: GATA3 acts as a pioneer factor in the cellular

process of mesenchymal-to-epithelial transition in breast cancer cells. GATA3 evicts histones from nucleosome-occupied chromatin in some regions and binds previously-inaccessible DNA and regulates transcription[48].

TF binding to chromatin is also regulated by motif sequence and motif location[51]. Both presence of TF binding motif and, distance, position and orientation of the motifs can determine consequence of TF binding. For example, IFN-β enhancer is 55 bp long and serves as the binding site for NFκB, IRF1, ATF2/c-Jun, CBP/p300 and architectural protein HMG I (Y)[59,60]. Binding of these factors on IFN-β enhancer forms a functional enhanceosome which regulate IFN-β gene transcription. However, changes in motif sequence of any of these TFs can drastically affect the binding capacity of all the TFs that have binding sites on IFN-β enhancer [51,56,61].

1.4.3 Nucleosomes in Transcription

The 2-meter long human DNA is fragmented into small pieces and is tightly packaged with nucleosomal proteins to form structured units called chromatin. This chromatin is then further condensed into stable organizational units called chromosomes. Humans have 46 chromosomes and they are all housed into an approximately 6 μm sized compartment in the cell called nuclei. Nucleosomes form the building blocks of chromatin and comprise four core histone proteins H2A, H2B, H3 and H4 and, histone H1 linker protein. Eight of these core histone proteins form an octamer around which 147 base pairs of DNA is wrapped[62]. Histone proteins have unstructured N-terminal tails that are heavily post-translationally modified. Some of the identified post-translation modifications (PTM) include acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ADP ribosylation, deimination, biotinylation, butyrylation, krotonylation, N-formylation, and

proline isomerization[62]. These PTMs give unique functional identity to the histones and combinatorial activity of these modified histones establishes the 'histone code'[63]. Based on these epigenetic 'codes', a gene can exist in different states such as active, inactive or poised. Histone H1 or linker histone binds DNA at its entry or exit point from nucleosomes and has also been found to harbor PTMs. These PTMs at histones serve as a major deterministic factor of chromatin states - whether regions of chromatin will be open and in a euchromatin region versus closed and part of a heterochromatin region. The process of dosage compensation and X-chromosome inactivation in mammals especially highlight the role of histones in switching off active chromatin regions into heterochromatin regions. X-chromosome inactivation is marked by the association of Xi-non coding RNA (Xist) as well as numerous repressive histone marks such as histone deacetylation, histone H3 lysine 4 (H3K4) demethylation, histone H3 lysine 9 dimethylation (H3K9me2), histone 3 lysine 27 trimethylation (H3K27me3), histone H4 lysine 20 monomethylation (H4K20me1), methylation of CpG islands at promoters, and enrichment of the histone variant macroH2A.

Histone modifications are highly dynamic and are continuously being monitored by a pool of 'writer' and 'eraser' proteins that either add or remove chemical modifications to histone tails and ensure proper histone functionality. During development, histone modifications at promoters and enhancers provide important directions in determination of cell differentiation. Aberrant histone methylation has been found to occur in various aging-related diseases including cancer. The ability of histones to respond to extracellular or intracellular signals give chromatin the flexibility to adapt and control gene expression, making histones a very interesting subject to study transcriptional dynamics[64,65,66].

1.5 Gene Regulatory Regions

1.5.1 Promoters

Promoters are the primary regulatory sites for transcription control and have been a widely studied topic. Promoter is located at the 5' end of the gene and can vary in length from 100bp to approximately 1000bp. The core promoter is a short approximately 100bp sequence and TSS is located near the center of this core promoter[67]. The core promoter contains 'core promoter elements' (CPEs)[68] and there are ~ 15 different promoter motifs that have been identified based on these CPEs: TATA-box, Inr, DPE, MTE, Ohler 1, Ohler 6, Ohler 7, DRE, TCT, BREu, BREd, DCE (I, II, III), XCPE1, XCPE2, Pause Button. Promoters differ in their motifs, transcription factors that bind to them and the strength or rate with which they can initiate transcription[67,69,70,71].

Among the different classes of promoters identified, TATA- containing promoters and TATA-less promoters are two major classes of promoters that have been extensively studied. TATA-binding proteins (TBPs) recognize the TATAA element, located 30bp upstream of the TSS on the core promoter, and bind to DNA along with other TBP-associated factors (TAFs)[71]. Transcriptional activators promote TBP binding to promoters that in turn leads to the assembly of pre-initiation complex (PIC) at promoters facilitating transcription. However, TATA box is present only in a minority (for example ~5% in fly) of core promoters but is very conserved from yeast to humans. The Initiator motif or Inr is more common than TATA-box but the sequence is not conserved across species. In TATA-less promoters, the downstream promoter element (DPE) is also present along with Inr and the precise position of these two elements is essential for TFIID binding to promoters and initiation of transcription[69].

Based on these core element motifs, pattern of transcription initiation, chromatin conformation and gene activity, promoters are divided into three categories[67] (Fig 1.4):

1. Promoters with sharp or focused initiation patterns, loosely organized nucleosomes mainly H3K4me3 – these promoters generally contain TATA-box and Inr motifs and are found at genes that are active in terminally differentiated cells.
2. Promoters with dispersed initiation patterns and precisely positioned nucleosomes (H3K4me3) with a defined nucleosome-depleted region - these promoters are generally present near or overlap a CpG island and are found at housekeeping genes.
3. Promoters with precisely positioned nucleosomes and bivalent histone marks such as H3K4me3 (active) and H3K27me3 (repressive) – these promoters typically show the characteristic of housekeeping promoters but are active in embryonic stem cells and are positioned at key development regulatory genes.

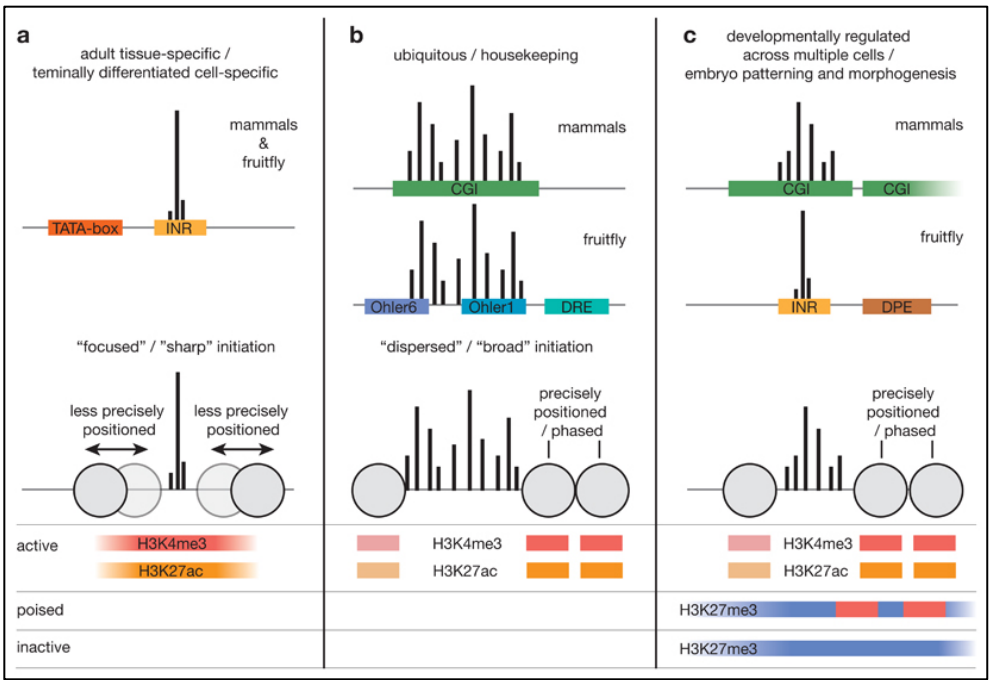


Figure 1. 4. Classification of major eukaryotic core promoters based on mode of transcription initiation and presence of histone marks. Figure reused from *Haberle and Stark, 2019, Nat Rev Mol Cell Biol. 2018 Oct; 19(10): 621–637* with permission

1.5.2 Enhancers

A cell is largely defined by the genes it transcribes[50]. These sets of genes distinguish one cell type from another. Studies show that along with promoters, enhancers also play important role in gene regulation. The classical definition of enhancers states that they are cis-regulatory DNA sequences that can be located either upstream, downstream or in the middle of a gene and irrespective of their distance and orientation, can regulate gene transcription[72]. However, with advancement in technology to identify enhancers, now it is being discovered how important enhancer location and orientation is in regulating gene expression. It has been estimated that there are ~1 million potential enhancers in the human genome making it almost 3% of the entire DNA sequence (assuming each enhancer length =100bp)[73]. This huge population of potential enhancers points to the fact that they have

been selected favorably though evolution and in turn raises the question of their functional relevance in regulating gene expression.

Enhancer was first discovered ~30 years ago in SV40 viruses as a 72bp sequence in late gene region of the virus[74]. The enhancer from SV40 was isolated and cloned near beta-globin gene in HeLa cells where it amplified expression of the gene 200-fold[74]. However, due to lack of proper sequencing tools genome-wide study of enhancers were impossible to carry out and enhancers remain poorly studied. With the recent emergence of global sequencing techniques such as ChIP-seq, PRO-seq, NET-seq and chromatin conformation studies such as Hi-C, the field of enhancer biology experienced a rapid growth and the profound influence of enhancers on gene regulation have been realized.

Recent studies show, unlike what was previously thought, enhancer 'grammar' (that is, the order, orientation, and spacing of binding sites of regulatory factors on enhancers) is critical to precise regulation of gene expression[75]. Farley *et al*., 2016 have shown that weak enhancers arranged optimally, and strong enhancers with strong-affinity binding sites for TFs arranged sub optimally can cause robust gene expression[76]. This precise mechanism of enhancer control through near placement of low TF binding affinity sites versus distal placement of high affinity TF binding sites is essential in regulating gene expression during development[76,77].

How enhancers regulate gene expression has always been an interesting question. There are several theories suggesting that enhancers act as the platform to bind transcription factors (TFs) that then recruit components of Mediator complex which together bring in RNA Pol II and mediate long-range interaction with target gene promoters[78]. This long-range enhancer-promoter interaction has been proposed to work in

two ways (a) either by looping[79] (b) or by tracking Pol II down the DNA and finally establishing a connection with promoters[78].

Recently it has been shown that enhancers can also be transcribed producing enhancer RNAs (eRNAs) the function of which is controversial[79,80,81,82,83,84]. Based on transcriptional activity and histone marks, an enhancer can be classified into three classes – inactive, poised and active[72].

(i) An inactive enhancer is generally characterized by presence of repressive histone marks such as H3K27me3 and Polycomb complex as well as lack of any detectable transcription.

(ii) A poised or primed enhancer is generally marked by bivalent histone marks such as H3K4me1 and H3K27me3. When hit with a stimulus, a poised enhancer switches to an active state and as a result enhancer RNAs are produced. Developmental enhancers and enhancers in stem cells are generally at a poised state.

(iii) An active enhancer is marked by eRNA transcription and presence of active histone marks such as H3K4me1 and H3K27ac, and binding of histone acetyltransferase such as CBP/p300.

These different states of enhancers make them flexible to respond to developmental as well as environmental stimuli making them important decision points of cell fate determination. Mutations in enhancer sequences have been found to occur in various developmental disorders and in diseases including cancer[85,86]. Looking at mutations in enhancer regions is a novel way of identifying underlying causes for diseases and this goes

against the traditional method of looking for mutations at only protein-coding genes in diseases.

Enhancers can be further categorized into shadow enhancers[3] and super enhancers[87]. Super-enhancers are large regulatory regions of DNA ranging from 10-20 kb in their length which is almost an order of magnitude higher than normal enhancers and act as a site to bind multiple TFs[88]. Super enhancers are also found to be associated with increased ES cell specific TF, and hence play a role in maintaining ESC state[89]. In cancers, super enhancers tend to be active near oncogenes such as MYC and FOS[90].

Shadow enhancers are similar to normal enhancers in structure and functions; they are named so as they reside further from the target gene compared to the primary enhancer or they are in the 'shadow' of the neighboring target gene. Recent studies show that these shadow enhancers are responsible for patterning and fine tuning of gene expression and play important role in evolution[91,92].

1.5.2.1 Enhancer RNAs.

Active enhancers marked by H3K4me1, H3K27ac, histone variant H2A.z, acetyltransferase p300, Mediator complex and shows open chromatin and DNase I hypersensitivity are capable of bringing in Pol II transcription machinery and initiating transcription[93]. eRNAs are generally short 1-2kb transcripts produced by bi-directional transcription from enhancer TSS[82]. The functions of these eRNAs are controversial. Studies have shown that eRNAs function in a variety of processes including regulation of transcription initiation from target gene promoters and acceleration of transcription by releasing NELFe from paused Pol II complex and promoting elongation. Another school of thought suggests eRNAs are simply transcriptional by-products. However, it is now

getting clear that eRNAs are involved in a diverse array of functions and some eRNAs clearly exert pronounced effect on mRNA transcription[94]. Contradicting previous notions that enhancers only accelerate gene expression, it has been shown that intragenic enhancers in mouse ESCs transcribe eRNAs that impede Pol II movement restricting transcription and inducing premature termination[95]. However, these intragenic enhancers can serve as extragenic enhancers for other genes accelerating their expression. This regulatory process between intra- and extra-genic enhancers serves as a powerful mechanism to control gene expression, especially during determination of cell lineage.

These observations clearly demand further studies of enhancers and eRNA in every cell type and determine the function of these small non-coding RNAs.

1.6 Research Objective

Through decades of research in the field, we have accumulated substantial insights into regulatory mechanisms of transcription and how these mechanisms change during rapid response to stimuli. However, every observation has led to the formation of a new question.

- How do cells maintain transcription when exposed to various intra and extracellular stimuli?
- Which genomic regions are more sensitive and show significant changes in response to these stimuli?
- How are histones affected during these rapid responses to stimuli?
- How does the dynamics of TF activity change during rapid response?
- Do all cells utilize the same mechanisms during rapid response?

- How rapidly does our chromatin conformation change when cells undergo rapid response?

In my dissertation thesis, I would be addressing some of these unanswered questions. In my second chapter, I would write about the methods I have used to address these questions. In my third chapter, I would characterize similarities and differences between heat-dependent and heat-independent stress responses in two different human cell types with a focus on the classical transcription factor HSF1, changes in chromatin conformation, and dynamics of promoter and enhancer-associated histone marks. In my fourth chapter, I would focus on the molecular mechanisms of Pol II transcription in the two cell lines exposed to the same stimulus. In my fifth chapter, I would use IL-1β as an inflammatory stimulus to elicit rapid transcriptional response in human cells and measure changes in genome-wide transcriptional landscape in a time-dependent manner to identify chromatin regions that respond rapidly to inflammation.

CHAPTER 2

METHODS

2.1 Cell Culture

MCF7 human breast adenocarcinoma were obtained from American Type Culture Collection (ATCC). Cells were cultured in DMEM/F12 (Life Technologies) media supplemented with 10% fetal bovine serum (FBS) (Gibco) at 37°C incubator with 5% CO_2.

Cells were treated with either heat-shock or Arsenic (Sigma). Heat-shock treatment was performed by rapidly placing a cell culture dish at 43°C water bath for 1 minute followed by incubation in a dry incubator for 1 hour at 43°C with 5% CO_2. Arsenic treatment was performed by incubating cells with media containing Arsenic (500μM) for 1 hour at 37°C with 5% CO_2. To avoid activation of HS -responsive genes due to addition of fresh media to the cells (Mahat 2016), fresh media was added 24 hours prior to the treatment.

A549 human lung adenocarcinoma cells were obtained from American Type Culture Collection (ATCC). Cells were cultured in RPMI-1640 (Life Technologies) media supplemented with 10% fetal bovine serum (FBS) (Gibco) at 37°C incubator with 5% CO_2. Cells were treated with 10ng IL-1β (Sigma – I 9041) for 15 minutes, 30 minutes, 45 minutes, 60 minutes, 120 minutes, and 5 hours, and compared with untreated cells.

NMuMG mouse mammary gland cells were obtained from American Type Culture Collection (ATCC). Cells were cultured in DMEM/F12 (Life Technologies) media supplemented with 10% fetal bovine serum (FBS) (Gibco) at 37°C incubator with 5% CO_2. They were used as spike-in controls for normalizing sequencing data from human cells.

2.2 Reverse Transcription Quantitative PCR (RT-qPCR) Gene Expression Analysis

Total RNA was extracted from cells using Qiashredder (Qiagen) and RNeasy kit (Qiagen) and checked for integrity using agarose gel electrophoresis. Total RNA concentration was measured using Nanodrop. 500ng -1μg of RNA was used to synthesize cDNA using random hexamer primers and SSRTIII reverse transcriptase enzyme (Life Technologies). Two-step real-time PCR was performed on these samples using primers self-validated and IDT-validated primers. Data were normalized against ACTB and GAPDH gene transcripts. Data were derived from at least two independent biological replicates and shown as mean +/- standard deviation. Statistical analysis was done using paired t-test (parametric) and one-way ANOVA with 95% confidence level using GraphPad Prism.

All primers for PCR and RT-qPCR reactions are listed in Table 2.1.

Table 2. 1. List of primers used for performing PCR reactions

Primer name	**Sequence**
KLF6 pre-mRNA F'	CTGGCAACAGGTAAACGCAGC
KLF6 pre-mRNA R'	AGGCCCAGCGGACCGCA
ACTB pre-mRNA F'	CGCCGCCAGGTAAGCCCG
ACTB pre-mRNA R'	GCGAAGGGGCCGCGGC
ZFN692_ChIP_F'	GTGAACTGGGGCAAGACTAA
ZFN692_ChIP_R'	GTAGCGCTGAGCTTCAACAC
ZNF692_up_f	GTTTCAGAGATGATATTCACA
ZNF692_up_r	CAGGTTGCCTATGGGAAGAAGT
ZNF692_dw_f	GGAGCAACCTGTCAGAAGAATTCAG
ZNF692_dw_r	CACTGTCTCAGCCACAGGAATCTG
SH3BP5L_ChIP_F'	GGCTCCGGGGGCAACTCGAG
SH3BP5L_ChIP_R'	AGTGGAGGGTTCCGCCGACT
SH3BP5L_UP_F'	AAGGGTGAAACACATTATCAG
SH3BP5L_UP_R'	CTGGGGGACAGTTTTGAGGC
HspA1BFL+0F	CCTGAGGAGCTGCTGCGACA
HspA1BFL+65R	AACGGGAGTCACTCTCGAAAAAG
HspA1BFL+10F	CTGAGGAGCTGCTGCGACA
HspA1BFL+183R	TTGGGGGCTGGAAACGGAA
Snail_ChIP_up_F'	ataattcttcacttcctctgggaa
Snail_ChIP_up_R'	ttctggtccagtgagggaga
Snail_ChIP_dw_F'	TGTTCCAGTCACAGCTGCTG
Snail_ChIP_dw_R'	GACGCCTTCATGGTACTCCT
hSnai1 + 29F	AGTGGTTCTTCTGCGCTACTGCT
hSnai1 + 139R	TCGCTGTAGTTAGGCTTCCGATTG

2.3 RNA-Sequencing

Total RNA was extracted from untreated cells and 5-hour IL-1β treated cells followed by RT-qPCR to validate expression of candidate genes using methods as previously described. Two independent biological replicates were sent to Novogene for library preparation and sequencing. Libraries were synthesized by Novogene with poly-A selection specifically enriching for stable RNAs and were sequenced in a paired-end manner with a sequencing depth of ~ 29 million reads (Table 2.2).

Table 2. 2. Table showing the RIN values of RNA samples that were sent for sequencing and their respective sequencing depth

Samples	Quality (measured by RNA Integrity Number or RIN)	Aligned reads
A549 IL1β 0min Rep#1	10	25784320
A549 IL1β 5h Rep#1	10	31947851
A549 IL1β 0min Rep#2	10	28998187
A549 IL1β 5h Rep#2	9.90	29442048

2.4 Precision Run-on Followed by Sequencing (PRO-Sequencing)

Nuclei isolation

Approximately 20 million cells were harvested, and nuclei were extracted using lysis buffer (10mM Tris-Cl, pH= 7.5, 2mM $MgCl_2$, 3mM $CaCl_2$, 0.5% IGEPAL, 10% glycerol, 2Units/ml Superase-In, Protease Inhibitor Cocktail). Spike-in mouse cells (5%) were added to human cells during nuclei extraction for all samples except one. Cells were pipetted gently in 3mL of lysis buffer for approximately 15 times using a P-1000 tip with the end cut off and incubated on ice for 5 minutes. Cells were centrifuged at 1000g for 10 minutes at 4°C, supernatant was discarded, and the process was repeated once more. Cell

pellet was then washed with 1mL of freezing buffer (50mM Tris-Cl, pH= 8.3, 40% Glycerol, 5mM $MgCl_2$, 0.1mM EDTA) and centrifuged using same parameters. Supernatant was discarded, and cell pellet was resuspended in ~60µl of freezing buffer making the total volume to 100µl and was either stored at -80°C or used for the next step.

Nuclear run-on reaction

In-vitro run on reaction was carried on at 37°C heated water-bath for 3 minutes using biotin-labelled nucleotides, 3µl each of biotin-11-ATP (PerkinElmer, cat. no. NEL544001EA), biotin-11-CTP (PerkinElmer, cat. no. NEL542001EA), biotin-11-GTP (PerkinElmer, cat. no. NEL545001EA), and biotin-11-UTP (PerkinElmer, cat. no. NEL543001EA) in a single tube. Total RNA was extracted by adding Trizol LS (Ambion; cat. no. 10296028) immediately to stop the run-on reaction and vortexed vigorously. Samples were incubated for 5 minutes at room temperature to allow complete dissociation of nucleoprotein complexes followed by chloroform (Sigma) extraction. Extracted RNA was then heat-denatured at 65°C for 40 seconds and immediately placed on ice.

RNA fragmentation

RNA is fragmented to a size of ~ 100bp by incubating the samples with 1M NaOH for 10 minutes on ice. Hydrolyzed samples were then purified using RNase-free P-30 buffer exchange column (Biorad).

First bead purification

Samples were enriched for biotin-labeled RNA by performing Streptavidin bead binding at room temperature for 20 minutes at ~1000 RPM. RNA-bound beads were then washed twice with ice cold high salt buffer, twice with binding buffer and once with low salt buffer. Bound-RNA was isolated from the beads using Trizol (Ambion) followed by

chloroform. Trizol extraction was performed twice on the same beads to maximize the amount of RNA obtained from the beads. The aqueous phase from both extractions were pulled together and RNA was precipitated by incubating with 100% ethanol and GlycoBlue at room temperature for 10 minutes followed by centrifugation at 14,000 RPM for 20 minutes at 4°C.

3’ adaptor ligation

1µl of 100µM 3’ RNA adaptor was diluted with 3µl of nuclease free water (NEB) and was used to dissolve RNA pellet from bead purification step. RNA was again heat-denatured at 65°C for 40 seconds and immediately placed on ice. 3’ adaptor ligation was carried at 20°C for 6 hours and incubated overnight at 4°C with T4 RNA ligase I enzyme (NEB), 10mM ATP, 50% PEG and RNase inhibitor Superase-In (Thermo Fisher).

Second bead purification

The samples were brought to 50µl with nuclease free water and a second Streptavidin purification for biotin-enriched RNA was performed in a similar fashion, as was previously described.

Decapping of 5’-capped RNAs

Bead purified RNA was dissolved in 5µl nuclease free water and heat denatured at 65°C for 40 seconds and immediately placed on ice. Pyrophosphates were removed from 5’ end of triphosphorylated biotinylated RNAs using RNA 5’ Pyrophosphohydrolase (RppH) enzyme (NEB) and incubating the reaction at 37°C for 1 hour.

5’-end phosphorylation

Decapped RNAs were then subjected to 5’-end phosphorylation by polynucleotide (T4-PNK) enzyme (NEB) with 10mM ATP and Superase-In at 37°C for 1 hour. RNA was then extracted using Trizol followed by triple chloroform purification. RNA was precipitated in 100% ethanol and GlycoBlue.

5’ adaptor ligation

1µl of 100µM 5’ RNA adaptor was diluted with 3µl of nuclease free water (NEB) and was used to dissolve 5’ end repaired RNA pellet. RNA was again heat-denatured at 65°C for 40 seconds and immediately placed on ice. 5’ adaptor ligation was carried at 20°C for 6 hours and incubated overnight at 4°C with T4 RNA ligase I enzyme (NEB), 10mM ATP, 50% PEG, and RNase inhibitor Superase-In (Thermo Fisher).

Third bead purification

The samples were brought to 50µl with nuclease free water and a third Streptavidin purification for biotin-enriched RNA was performed in a similar fashion, as was previously described

Reverse transcription

RNA pellet was dissolved in 10µl nuclease free water and cDNA was synthesized using 100µM RP1 reverse transcription primer at a final concentration of 2.5 µM and 625 µM dNTP mix, 0.1M DTT, Superase-In, and reverse transcriptase SSRT III. The mixture was incubated at 45°C for 15 minutes, at 50°C for 40 minutes, at 55°C for 10 minutes, and at 70°C for 15 minutes. Total volume of samples was brought to 26µl.

PCR amplification

PCR amplification of cDNA templates were performed in real-time using 250 nM of each RP1 and RPI-1 primer (IDT), 250 µM dNTP mix, 1x HF buffer, 1M Betaine, 1x Evagreen dye, and Phusion DNA polymerase. Amplification curves were observed in real-time on a Biorad quantitative-PCR machine and reaction was terminated at 1500-2000 RFU x 10^3 or relative fluorescence units which is reached approximately in 10 to 13 cycles, depending on the samples.

In some cases, full-scale PCR amplification cycle numbers were determined by taking 2µl of cDNA sample and making dilutions corresponding to 17, 15, 13, and 11 full-scale PCR cycles. Amplified test cDNA was run on 2.2% agarose gel with Blue DNA loading dye in 1X TAE buffer at 100V for 15 minutes and 130V for 45 minutes. The gel was stained with either 10µl of Ethidium Bromide or 15µl of SYBR Gold in 150 ml of 1X TAE buffer. The gel was stained for 10 minutes and imaged with UV-light. Cycle numbers for full-scale amplification were determined where there was sufficient amount of product with 50 -75% unused primers (Fig 2.1).

Library size selection

PCR-amplified products were then run on 6% TBE gel (Novex) with 1X Tris-Borate-EDTA (TBE) buffer with 100bp DNA ladder (NEB) at 120V for 45 minutes. The gel is stained with 8µl ethidium bromide in 100ml 1X TBE for 5 minutes with gentle shaking. Stained gel is then imaged with UV-light and bands from ~ 125bp-300bp are excised with a clean razor (Fig 2.2). Gel fragments are then sheared in gel-braker tubes placed in 2 ml eppendorfs. Gel fragments are then dissolved in 400 µl 1X Tris-EDTA (TE) buffer (Sigma) by vigorous shaking at room temperature for 3 hours. Dissolved gel

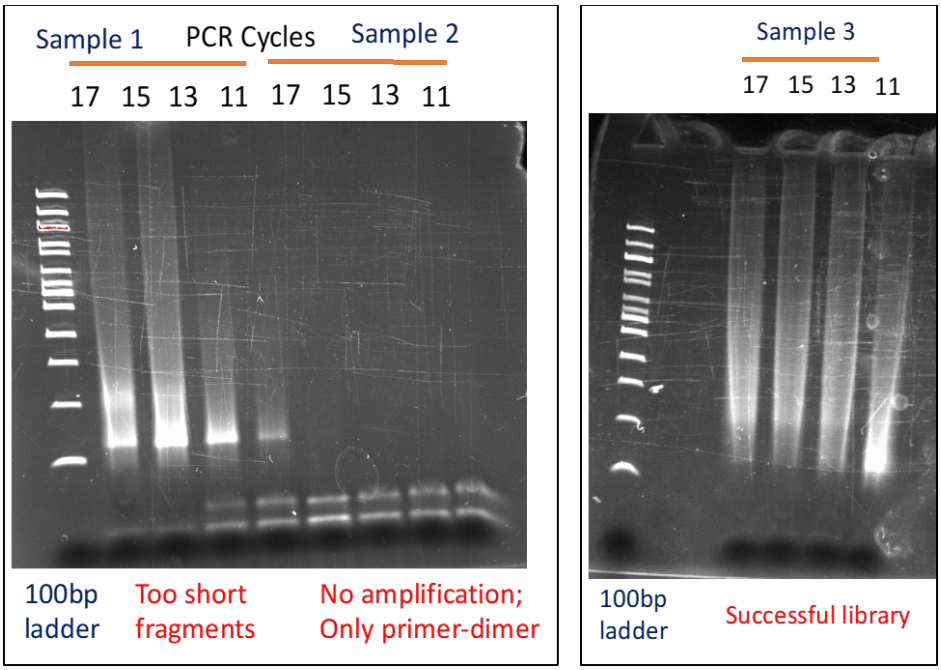


Figure 2. 1. Test amplification of PRO-seq samples. 2µl cDNA samples amplified using different PCR cycles (17, 15, 13, 11) run on 6% Tris-Borate-EDTA (TBE) non-denaturing gels to determine the minimum number of PCR cycles for amplification of total library. Left: TBE gel with unsuccessful libraries showing either too short fragments or no enrichment at all. Only unused primers can be seen at the bottom. Right: A successful test amplification run on a TBE gel.

fragments are then purified through cellulose-acetate spin filters (Agilent) to get rid of any undesired particles. Library is then extracted with 300 mM sodium acetate (Sigma), 1 µl GlycoBlue, and 100% ethanol.

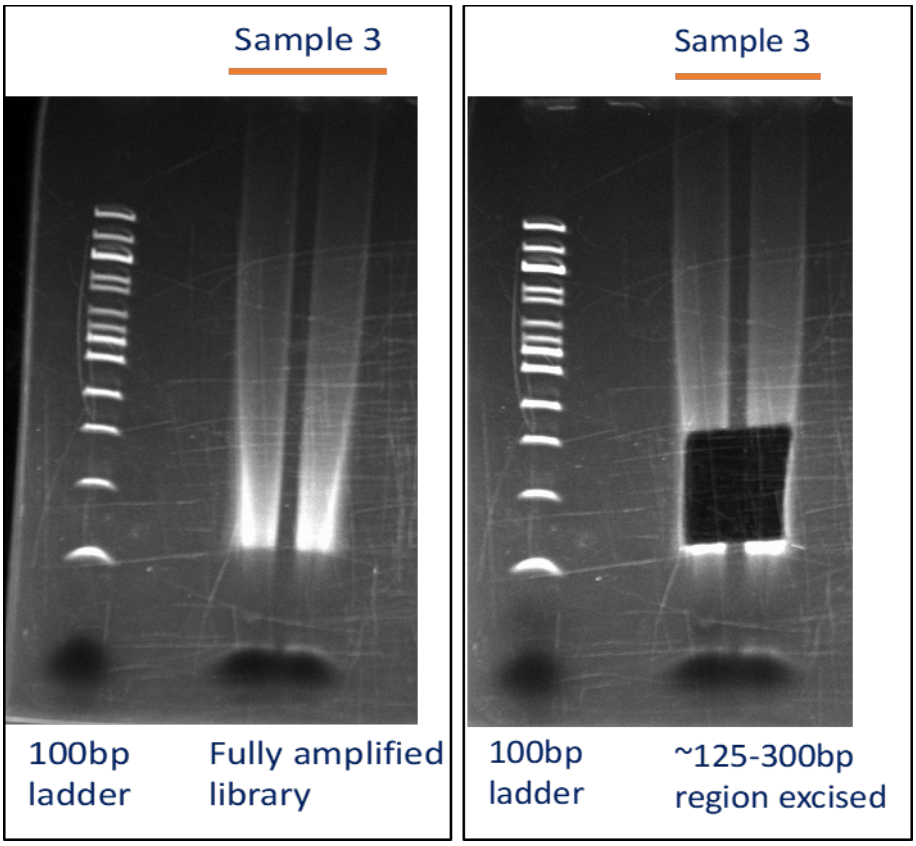


Figure 2. 2. Size selection by gel purification of PRO-seq library. (Left) Fully amplified library (after PCR cycle number determination) run on 6% TBE non- denaturing gel. (Right) Size selection using gel purification

Sequencing

Samples were first quality-controlled on a BioAnalyzer and fragment sizes from libraries were validated to be within the range of expectation (Fig 2.3).

They were further quality-controlled by running on a MiSeq sequencer generating ~ 3 million paired-end reads. If the sample showed biological difference among different time points of treatments and similarity among replicates, they were sent to Novogene for high throughput single-end sequencing from 3' end generating ~ 25-30 million reads per sample.

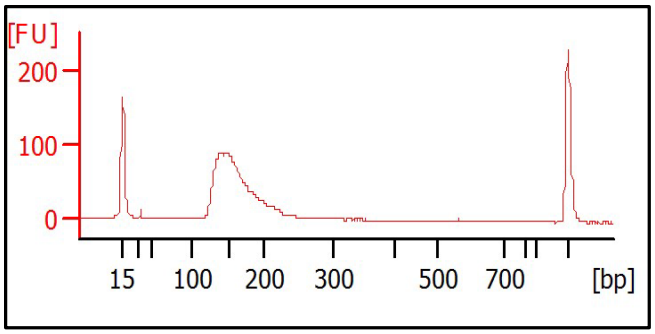


Figure 2. 3. Bioanalyzer image of a successful PRO-seq library showing peak around ~150-200bp

2.5 Chromatin Immunoprecipitation Followed by Sequencing (ChIP-seq)

Crosslinking

Approximately $2*10^7$ cells were crosslinked with 1% methanol-free formaldehyde (16% formaldehyde solution w/v from Thermo Scientific) in serum-free DMEM/F12 media rotating in a slow shaker at room temperature for 10min followed by quenching with 0.125M Glycine. Cells were washed with 1x phosphate-buffered saline (PBS) thrice and lysed by incubating on ice with lysis buffer with 1% sodium-dodecyl sulfate (SDS) (ingredients) for 15min.

Sonication

Sonication condition was first standardized using different time points (0min, 2min, 4min, 6min, and 8min) in a Covaris sonicator with 200 cycles per burst, duty factor of 5 and peak power of 140. Ideal sonication condition was determined by running sonicated samples on 1.5% agarose gel and performing western blot against the protein-of-interest (Fig 2.4).

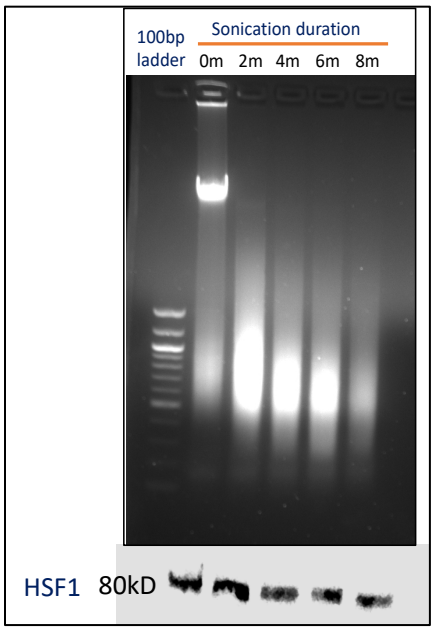


Figure 2. 4. DNA was extracted from sonicated samples and were run on a 1.5% agarose gel with an unsheared control sample (0 min) with 100bp DNA ladder to check size of shearing. From the same sonicated samples, protein was extracted and protein-of-interest (eg: HSF1) for ChIP was immunoblotted to check integrity (bottom panel).

Following standardization, all samples were sonicated for 6 min in Covaris sonicator using 200 cycles per burst, duty factor of 5 and peak power of 140. Sonicated cells were centrifuged at 10,000 RPM for 10 minutes at 4°C to get rid of any debris and the supernatant was collected. 2% of the supernatant was collected and validated for sonication of chromatin by reverse-crosslinking and extracting DNA and running it on a 1.5% agarose gel stained with ethidium bromide at 150V with 100bp ladder (NEB). Another 2% of supernatant was collected for validation of protein integrity by performing Western Blot against candidate proteins.

Pre-clearance of sonicated samples

Magnetic beads A+ G in the ratio 1:1 were washed twice and resuspended in dilution buffer. Sonicated samples are pre-cleared with washed magnetic beads A + G for 1 hour in a slow rotor at 4°C. Beads were removed by placing the tubes on a magnetic stand and 1% of the sample were taken as inputs. At least two inputs were taken for each sample.

Antibody binding

Precleared samples were incubated with 5µg of each of the antibodies listed below (Table 2.3) in 2ml dilution buffer (1% Triton-X, 2mM EDTA, 20mM Tris pH=8.0, 150mM NaCl, 1X PMSF, protease inhibitor) by slow rotation overnight at 4°C.

Table 2. 3. List of antibodies used for performing ChIP experiments

Primary Antibody	**Company purchased from**	**Dilution used**
H3K4me1	Abcam	5µg
H3K4me3	Abcam	5µg
Pan H3	Abcam	5µg
H3K27ac	Abcam	5µg
RNA Pol II CTD	CST	5µg
HSF1	Enzo	5µg

Magnetic bead purification

Antibody-bound samples were enriched by incubating with prewashed magnetic beads for 2 hours at 4°C in slow rotor. Beads were separated using magnetic stand and supernatant containing unbound antibody or chromatin was discarded. Beads were washed

once with ice-cold low salt buffer, thrice with ice-cold high salt buffer, once with lithium chloride buffer, and twice with TE buffer by slow rotation for 5 minutes at room temperature. Beads were resuspended in SDS elution buffer, and chromatin was extracted by shaking the solution at 900 RPM for 30 minutes at 65°C.

DNA Extraction

Chromatin is reverse-crosslinked by incubating the samples overnight in 200mM sodium chloride at 65°C. Samples were then treated with Proteinase-K for 2 hours at 56°C in presence of Tris-Cl (pH=8) (Sigma) followed by phenol-chloroform (Sigma) extraction of DNA. DNA was precipitated using 300mM sodium acetate, 1µl GlycoBlue, and 100% ethanol. DNA was resuspended in 1x TE buffer and real-time qPCR was performed for positive and negative control primers as listed in Table 2.1.

Library preparation

ChIP DNA library was prepared using qPCR-validated samples following NEBNext Ultra II DNA library prep kit (NEB) protocol without performing size selection. Library was validated using qPCR against positive and negative control primers. Validated libraries were sent for further quality-control with Bioanalyzer and sequencing (Fig 2.5).

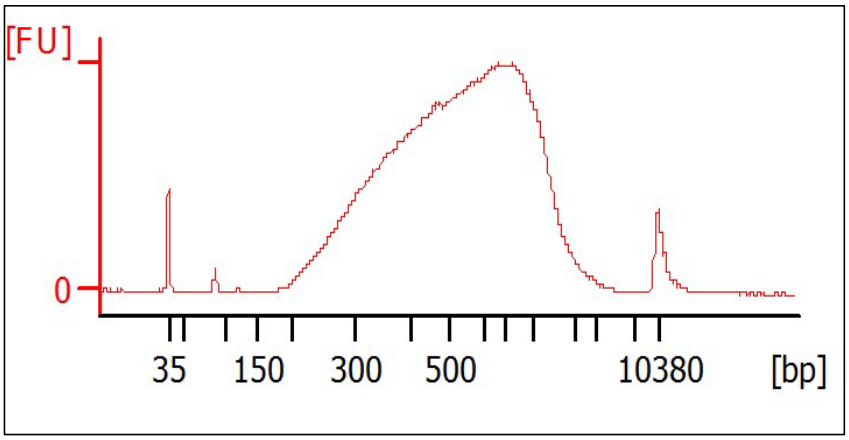


Figure 2. 5. ChIP sequencing library run on High-Sensitivity Bioanalyzer Chip for detecting the range of fragment distribution and as a quality-control step before running on MiSeq

2.6 ATAC-Sequencing

Cell harvesting

Cells were grown in 10cm dishes with complete media supplemented with 5% FBS for a day. Cells were harvested by incubating with 2ml of 0.25% of Trypsin for 2-4 minutes. Complete media was added to trypsinized cells immediately to stop the reaction and were centrifuged at 500 g for 5 minutes. Cell pellet was collected and washed with cold 1X PBS by centrifugation.

Cell Lysis

Cells were then resuspended in cold 1X PBS at a concentration of 1x 10^6 cells/mL. 5 x 10^3 cells (~50 µl) were collected and centrifuged. Supernatant was discarded and cells were lysed using equal volume (~50 µl) of Greenleaf's lysis buffer (Tris ph=7.5, $MgCl_2$, NaCl, 0.5% IGEPAL) for 5 minutes on ice. Two parallel tubes of cells were lysed together using the same protocol – one was used to cell count and rupturing of nuclei after lysis, the

other one was used for further processing of ATAC. After lysis, the cells were spun at low speed 500 g for 5 minutes at 4°C.

Transposase reaction and PCR amplification

Supernatant was discarded and 50µl transposase reaction mix (2x TD buffer = 25µl, Tn5 Transposase = 2.5 µl, Nuclease free water = 22.5 µl) was added. Cells were continuously shaken at a low speed of 400 RPM at 37°C for 30 minutes. DNA was purified using MinElute Qiagen kit and amplified using NEB Nextera kit using our modified real-time PCR protocol.

2.7 Western Blot

Cells were lysed in Urea lysis buffer (8M Urea, 1% SDS, 126mM Tris pH 6.8) and protein concentration was measured in Qubit 1.0 fluorometer using Qubit protein assay kit (Invitrogen). Cell lysates corresponding to ~ 30ug of protein were run on a 10% SDS-PAGE resolving gel (BioRad) at 200V for 1 hour followed by transfer of protein from gel unto a polyvinyl fluoride or PVDF membrane in 1x transfer buffer (100ml of 10X Tris-Glycine-SDS and 200ml of methanol in 700ml Miili-Q water) for 2 hours at a constant current of 100V at 4°C. Non-specific antibody binding sites were blocked by incubating the blots in 5% milk diluted in 1x TBS with 0.1% tween (TBS-T) for 30 minutes. Blots were washed thrice for 5 minutes with vigorous shaking at room temperature and incubated with primary antibody as listed in Table 2.4 at concentration validated by the manufacturer or in our lab overnight at 4°C on a slow rotator. The blots were washed with 1x TBS-T three times for 10 minutes each. The blots were incubated with secondary antibody conjugated to horse-radish peroxidase (HRP) (GE) at a dilution of 1:10,000 or 1:5000 respectively for 1 hour at room temperature on a slow rotator. Blots were washed with 1x

TBST three times for ten minutes each and either imaged directly on an Odyssey imager using the appropriate absorbance spectrum of the respective fluorophore or incubated with chemiluminescent HRP substrate (Amersham ECL GE) and imaged on an Odyssey imager, respectively.

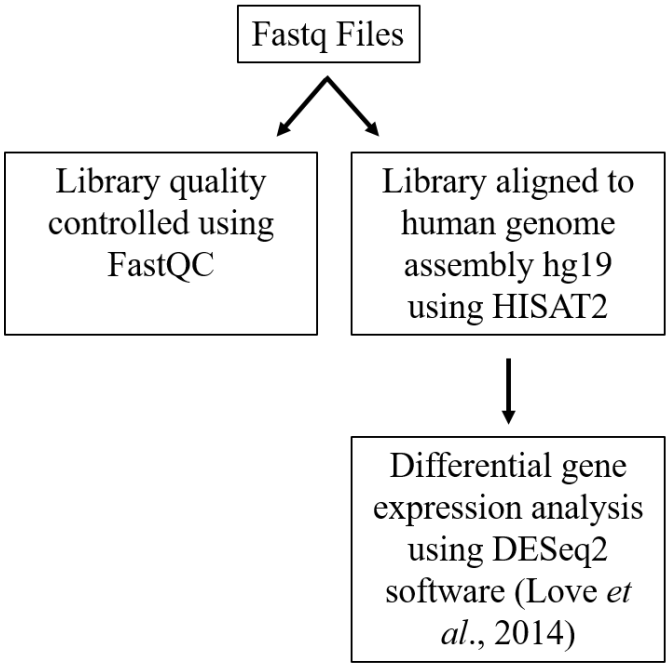
Table 2. 4. List of antibodies used for performing Western Blot experiments

Primary Antibody	**Company purchased from**	**Dilution used**	**Secondary antibody**	**Company purchased from**
Lyn	Santa Cruz: Cat#sc-15; Lot#D2012	1:1000	Anti-rabbit HRP	GE
Phospho-Lyn	Santa Cruz: Cat#sc-03; Lot#A0516	1:1000	Anti-rabbit HRP	GE
NFκB/p-65	Santa Cruz: Cat#sc-8008; Lot#K0513	1:1000	Anti-mouse HRP	GE
β-Actin	Miilipore; Cat#MAB-1501; Lot#2604822	1:2500	Anti-mouse HRP	GE
HSF1	Enzo	1:1000	Anti-rabbit HRP	GE
p-HSF1	Enzo	1:1000	Anti-rabbit HRP	GE

2.8 Bioinformatic Analysis

2.8.1. RNA-Seq Analysis

Libraries were aligned against hg19 human genome with HISAT2 alignment tool. Libraries were converted to BAM format using SAM tools and differential gene expression was analyzed with DESeq2 program performed on R with a p value of 0.01.



2.9 ChIP-Seq Analysis

Libraries were aligned to human hg19 genome using HISAT2 and bam files were generated. Bam files were converted to BigWig files using bamCoverage with the default bin size of 50bp. These files were uploaded to UCSC Genome Browser for visualization.

Duplicate reads from aligned bam files were removed using picard tools 'mark duplicates' where duplicate reads are defined as arising from the same DNA fragment with same 5' end. After detecting the duplicate reads, picard assigns scores to each of the reads based on the sum of their base quality and differentiates between primary and duplicated reads.

Peaks were called independently in each sample using MACS2 peak calling software with a q-value or FDR cut-off of 0.01 selecting for only strong peaks. Furthermore, only those peaks are selected for further analysis which were found to be

present in both replicates of the library. However, for some analysis, all called peaks from all samples were used.

Heatmaps and metaplots were generated using deepTools plotHeatmap and plotMetaplot function with BigWig files normalized internally using Counts Per Million (CPM) reads.

MACS2 narrow peak files were converted into genomic ranges (Granges) object and peaks were annotated using BioConductor package 'ChIPseeker' in R. Overlap of peaks among replicates were also performed using 'ChIPseeker' package function 'findOverlapofPeaks' with a maximum distance of 1000bp.

Motif analysis was done using HOMER tool 'findMotifsGenome' and size of region specified for motif search was 200, which was also the recommended size for motif analysis by HOMER.

2.10 ATAC-Seq Analysis

Adapters were trimmed from fastq reads of ATAC libraries using 'trimmomatic'. Reads were then aligned to human hg19 genome using HISAT2 and bam files were generated. Bam files were converted to BigWig files using bamCoverage with the default bin size of 50bp. These files were uploaded to UCSC Genome Browser for visualization.

Duplicate reads from aligned bam files were removed using picard tools 'mark duplicates' where duplicate reads are defined as arising from the same DNA fragment with same 5' end. After detecting the duplicate reads, picard assigns scores to each of the reads based on the sum of their base quality and differentiates between primary and duplicated reads.

Peaks were called independently in each sample using MACS2 peak calling software with a q-value or FDR cut-off of 0.01 selecting for only strong peaks. Furthermore, only those peaks are selected for further analysis which were found to be present in both replicates of the library. However, for some analysis, all called peaks from all samples were used.

Heatmaps and metaplots were generated using deepTools plotHeatmap and plotMetaplot function with BigWig files normalized internally using Counts Per Million (CPM) reads.

MACS2 narrow peak files were converted into genomic ranges (Granges) object and peaks were annotated using BioConductor package 'ChIPseeker' in R. Overlap of peaks among replicates were also performed using 'ChIPseeker' package function 'findOverlapofPeaks' with a maximum distance of 1000bp.

2.11 PRO-Seq Analysis

Adapters (sequence = TGGAATTCTCGGGTGCCAAGG) were removed from PRO-seq libraries using 'cutadapt' and reads shorter than 15 bp in length were discarded. Trimmed reads were aligned against human ribosomal RNA (rRNA) to get rid of rRNA contamination of nascent transcripts and filtered reads were used for further processing. For generating rRNA index assembly, 42,999 bp linear human ribosomal DNA sequence was obtained from GenBank: U13369.1, NCBI and indexed using 'bowtie-build'. Unaligned rRNA reads were now aligned against human hg19 using HISAT2 and BAM files were generated. BigWig files were generated from BAM files and uploaded on UCSC Browser for visualization.

These BigWig files were normalized using multiple ways to rule out technical effects from the data. The files were normalized against –

a) read depth of spike-in control from mouse nuclei
b) 3’ end of long genes where Pol II could not have reached within the time of treatment if it progressed at 3kb/min. If treatments were 60 minutes long, the region of genes selected for normalization was from (60 minutes *3kb) 180kb plus 20kb buffer region to 0.5kb before the gene ends to avoid the polyA tail.
c) Counts per million (CPM)

Heatmaps and metaplots were generated using deepTools plotHeatmap and plotMetaplot function with normalized BigWig files.

Differential changes in gene expression was done using DESeq2 with a p-value cut-off of 0.05 on gene bodies. Gene body was defined as TSS +200bp to +120kb.

Pathway analyses of differentially regulated genes were done using Bioconductor packages ‘ChIPseeker’ and ‘clusterProfiler’.

CHAPTER 3

GLOBAL HEAT SHOCK RESPONSE IS CELL-TYPE SPECIFIC AND REVEALS DISTINCT CHANGES AT INTERGENIC REGIONS OF CELLS, INDEPENDENT OF HEAT STRESS

3.1 The Heat Shock Response (HSR)

Life has been found to prosper in a diverse range of temperatures from below freezing point to as high as 113°C (Stetter et al, 2006)[96]. Temperature plays a huge role in sustenance of life including growth and reproduction. Even minor changes in temperature can have huge impact on the survival of all living organisms (Stetter et al, 2006)[96]. Changes in temperature such as exposure to high heat can cause denaturation of proteins, abnormal protein folding, accumulation of abnormal proteins resulting in proteotoxic stress and loss of cellular homeostasis in living organisms [97]. In eukaryotic cells, heat stress not only affects individual proteins but leads to major changes in subcellular components including actin filaments, endoplasmic reticulum, mitochondrial structures and lysosomal compartments within the cell [97] (Richter, 2010). Thus, even slight increases in temperature can lead to death of organisms. Therefore, cells have evolved a very conserved pathway of buffering stress response called the heat-shock response or HSR pathway. HSR has been found in plants, bacteria, yeast, slime molds, fishes, sea-urchins, flies, and mammals. HSR is thus a highly conserved, probably universal mechanism present in most of the organisms that have been studied till date[98]. The temperature at which an organism would mount HSR

is specific to the organism itself. For example, E. coli shows HSR at 42°C while HSR in salmon is seen at 24°C [98].

The effect of heat stress has been shown as early as in 1930s in fruit fly Drosophila (Goldschmidt, 1935; cf. Morimoto et al., 1990; Stetter 2006). Flies are normally grown at room temperature around 25°C [98] (Lindquist, 1986), however when temperature is raised to 29°C flies show HSR, the maximum effect being seen between 36°C and 37°C [98] (Lindquist, 1986). At 37°C, Drosophila melanogaster show rapid changes in chromosomal structures and associated transcription[99] (Jamrich, 1977). Active regions of transcription in chromosomal puffs of Drosophila salivary gland lose RNA Polymerase stain [99,100] (Ritossa, 1962; Jamrich, 1977), whereas new sites of transcription appears in the puffs as shown in the figure below [99] (Jamrich, 1977) (Fig 3.1)

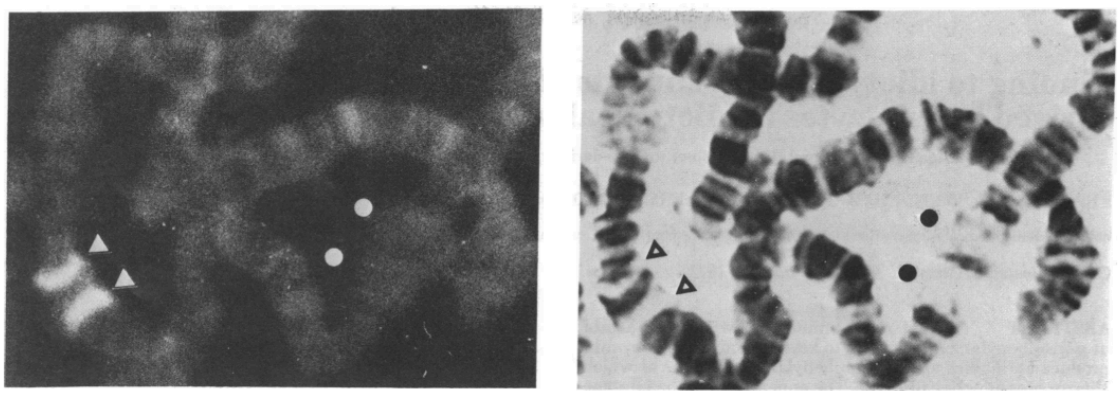


Figure 3. 1 Immunofluorescent images showing polymerases disappear from puffs in *Drosophila* after heat-shock for 45min. Arrows indicate appearance of RNA Pol II after heat-shock whereas circles indicate regions from where polymerases are removed as a result of heat-shock. Figure reused from *Jamrich et al., PNAS May 1, 1977 74 (5) 2079-2083* with permission.

Studies show that there are approximately seven functional groups of proteins that are upregulated in HSR (Fig 3.2)[97]. However, the members of these groups and the extent of upregulation is highly variable among organisms. These group of proteins include:

1. Hsps or heat-shock proteins such as hsp70, hsp90 that act as molecular chaperones
2. Proteins involved in proteolysis to clear misfolded and irreversibly aggregated proteins
3. DNA/ RNA repair proteins to correct double-stranded nucleic acid breaks or epigenetic mutations such as aberrant methylations
4. Metabolic enzymes to balance energy supply
5. Transcription factors and kinases involved in regulating the response itself
6. Proteins involved in protecting cytoskeletal structures and cell organization
7. Membrane transport and detoxifying proteins

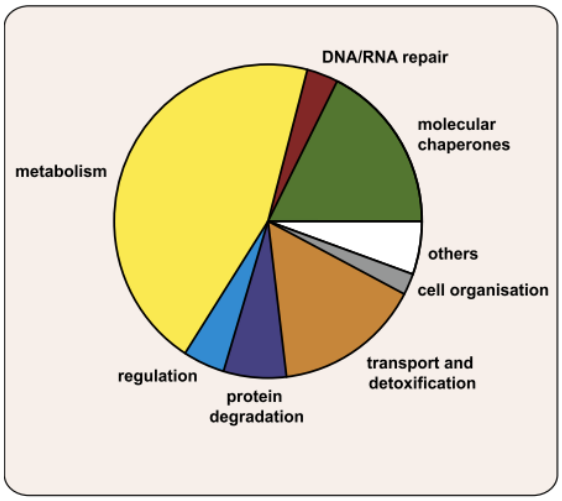


Figure 3. 2. Summary of the functional classes of genes upregulated during the heat shock response in S. *cerevisiae* after an increase in temperature from 25°C to 35°C for 10 min. *Figure taken from Richter et al., Molecular Cell; Volume 40, Issue 222 October 2010Pages 253-266 with permission.*[97]

3.2 Heat Shock Factors

In eukaryotes, heat-shock response proteins are regulated at transcriptional level by a class of transcription factors (TF) called heat-shock factors (HSFs) (Wu, 1984) [98]. Till date, a single HSF has been identified in invertebrates while four HSFs: HSF1, HSF2, HSF3 and HSF4 has been identified in vertebrates [101]. Among the heat shock factors studied, HSF1 is the most well characterized and is found to be the chief activating factor for HSR genes such as hsp70 in heat shocked cells. HSF1 is constitutively expressed in cells at physiological temperature and is present mostly in the cytosol as an inactive monomer. HSF1 forms a complex with hsp70 or hsp90 chaperone proteins which block it from being released in unstressed cells. With heat-shock, hsp70 and hsp90 bind denatured proteins and release HSF1[102]. HSF1 gets heavily phosphorylated, homo-trimerized and transported to the nucleus where it then binds to heat-shock response elements (HSE) present on DNA.[103] HSE consists of 3–6 alternating, inverted pentameric repeats (nGAAn). HSF1 binds HSEs through a well conserved winged helix-turn-helix DNA binding domain [101]. HSF1 itself has heat-sensing capacity as it can spontaneously trimerize in-vitro with heat stress[101,104,105,106]. Below is a schematic of HSF1 functional domains with their respective post-translation modifications from Åkerfelt et al., 2010[101] (Fig 3.3).

HSF1 protein is composed of 529 amino acids which undergoes several post-transcriptional modifications (Fig 3.3). At N-terminal of the protein lies the DNA binding domain (DBD) which undergoes acetylation (A) by p300. Acetylation of the DBD decreases HSF1's ability to bind to DNA whereas deacetylation by sirtuin SIRT1 increases its DNA binding ability. The heptad repeats (HR-A and HR-B) are hydrophobic repeats that mediate trimerization of HSF1 forming a coiled-coil structure whereas and HR-C

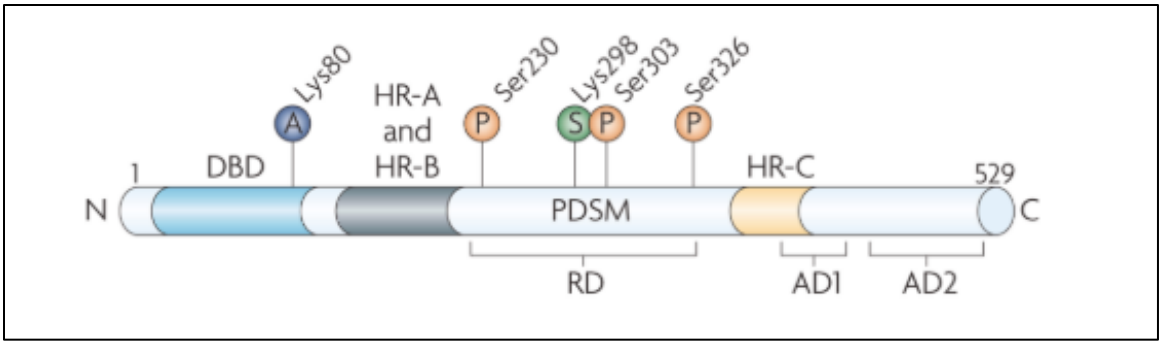


Figure 3. 3. Schematic diagram of HSF1 protein showing its different binding domains. Figure reused from *Åkerfelt et al., Nat Rev Mol Cell Biol, 11 (8), 545-5; 2010 with permission*

suppresses spontaneous trimerization of HSF1. The regulatory domain (RD) contains phosphorylation-dependent sumoylation motif (PDSM) and undergoes rapid phosphorylation (P) and sumoylation (S) upon heat-shock leading to transactivation of HSF1. The transactivation domain is composed of two modules — AD1 and AD2, rich in hydrophobic and acidic residues, and is responsible for rapid stress response upon.

HSF1 is activated not only during heat shock, but also in response to heavy metal stress such as arsenic, copper, zinc, cadmium, and silver[107]. HSF1 is also shown to be activated by anti-inflammatory drugs such as Celastrol[108]. These observations highlight the fact that HSR pathway can be activated by both heat-dependent stress and heat-independent stress. It has been recently shown in mouse cells that HSR leads to HSF1 binding at and outside of promoters[103]. However, if HSF1-mediated HSR is similar or different between heat-dependent and heat-independent stimuli in the same system in terms of HSF1 binding, transcriptional output and chromatin accessibility have not been looked at. In my study I have dissected the mechanisms of transcriptional responses in two different human cell types exposed to heat stress and heat-independent heavy metal stress both of which rapidly induce HSR. We also asked if how two human cell lines, of very different origin, both

activating HSR pathway in response to heat shock utilize a similar mechanism of transcriptional activation upon heat stress.

3.3 Motivation

Because of its importance in basic science as a model of transcription activation and its health relevance in diseases such as cancer, HSR has been studied extensively. HSR has been shown to be induced by heat-dependent and heat-independent stimuli in normal cells as well as HSR can be induced by cancer cells as a protective mechanism to survive in the body. We asked whether cells utilize the same program of HSR when exposed to temperature-dependent and temperature-independent stresses. We further questioned if HSR follows the same program in two different cell types. In this section, I would elucidate my findings of HSR across different cell types exposed to heat-dependent and heat-independent stressors.

3.4 MCF7 Cells Elicit Robust HSR When Exposed to Heat-Shock (HS) and this Response is Transient

MCF7 cells were subjected to rapid HS at 43°C (this is a standard HS temperature that has been used by several labs through decades) in a time-dependent manner – 0 minutes, 15 minutes, 40 minutes, 60 minutes, and 180 minutes, and changes in gene expression was measured with RT-qPCR against known heat-shock responsive genes such as HSPA1B (hsp70), and HSPH1. (Fig 3.4)

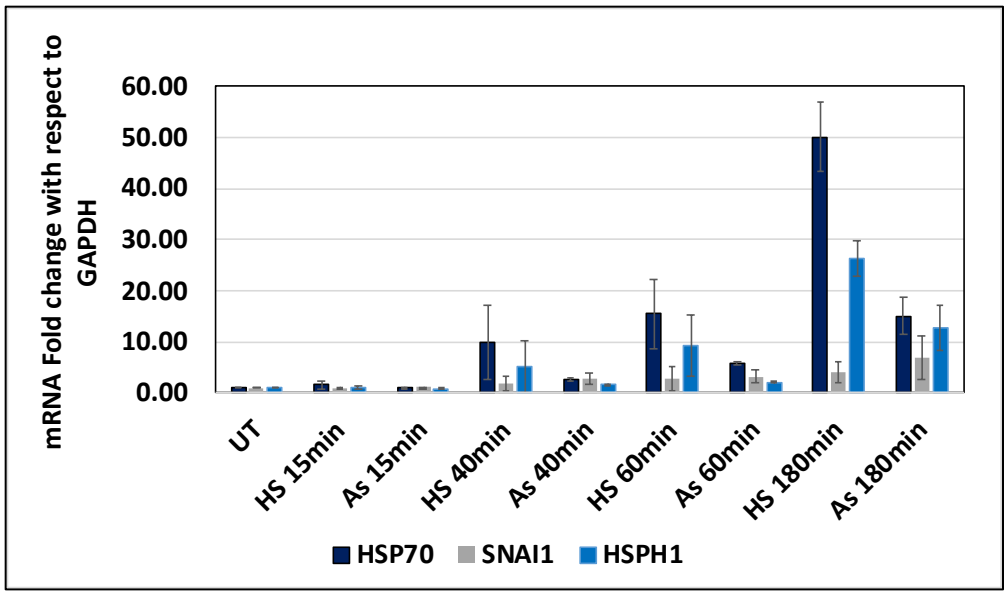


Figure 3. 4. Time-dependent HSR induced by heat-dependent and heat-independent stresses. RT-qPCR on heat shock and arsenic treated MCF7 cells (n = 2, $p<0.01$) over a time-course showing upregulation of transcription in heat-shock responsive genes. Statistical analysis is done using non-parametric unpaired t-test followed by Mann-Whitney test.

To determine if exposure to such high temperature causes temporary or permanent change in expression of HS responsive genes in MCF7 cells, a HS pulse experiment was performed where cells were heat-shocked for 60 minutes at 43°C and then kept at 37°C for 4 hours. RNA was extracted from these cells and RT-qPCR was performed showing that HSR to temperature is transient and normal gene expression level is restored (Fig 3.5).

3.5 Standardization of ChIP-Sequencing (ChIP-seq) Technique

Before proceeding to the major ChIP-seq experiments, we first established the protocol of ChIP-seq using RNA Pol II CST antibody. A 1:1 mixture of magnetic Protein A and magnetic Protein G beads were used to chromatin bound to Pol II or chromatin without Pol II. Increasing concentrations of beads: 1X (15 µl), 2X (30 µl) and 4X (60 µl) were used for the same number of cells (~ 15 million) and same amount of Pol II.

antibody (5 µl). Increasing concentration of beads led to increased pulldown of protein-bound chromatin but simultaneously decreased signal-to-noise (S/N) ratio. 2X concentration of beads were decided to be ideal for further ChIP-seq experiments (Table 3.1 and Fig 3.6).

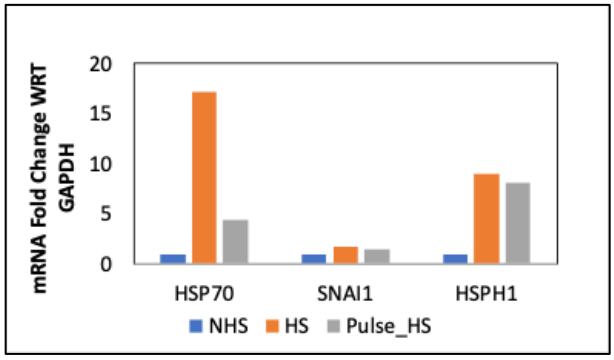


Figure 3. 5. HSR is transient. RT-qPCR (n=1) of HS-responsive genes in MCF7 cells showing upregulation with heat-shock and return to near basal state after removal of heat-stress indicating HSR is rapid and transient.

Table 3. 1. Table showing signal-to-noice ratio with addition of different concentrations of Dynabeads protein A and Dynabeads protein G.

Samples	Primers	% input	S/N
IP_+P2_1X	Snail Up	0.0011	
	Snail Promoter	0.1518	141.2
	Snail down	0.0014	
IP_-P2_1X	Snail Up	0.0008	
	Snail Promoter	0.0028	3.507
	Snail down	0.0002	
IP_+P2_2X	Snail Up	0.0043	
	Snail Promoter	0.4648	107.3
	Snail down	0.0028	
IP_-P2_2X	Snail Up	0.0006	
	Snail Promoter	0.0028	4.611
	Snail down	0.0005	
IP_+P2_4X	Snail Up	0.0069	
	Snail Promoter	0.6182	89.62
	Snail down	0.006	
IP_-P2_4X	Snail Up	0.0025	
	Snail Promoter	0.0064	2.571
	Snail down	0.0014	

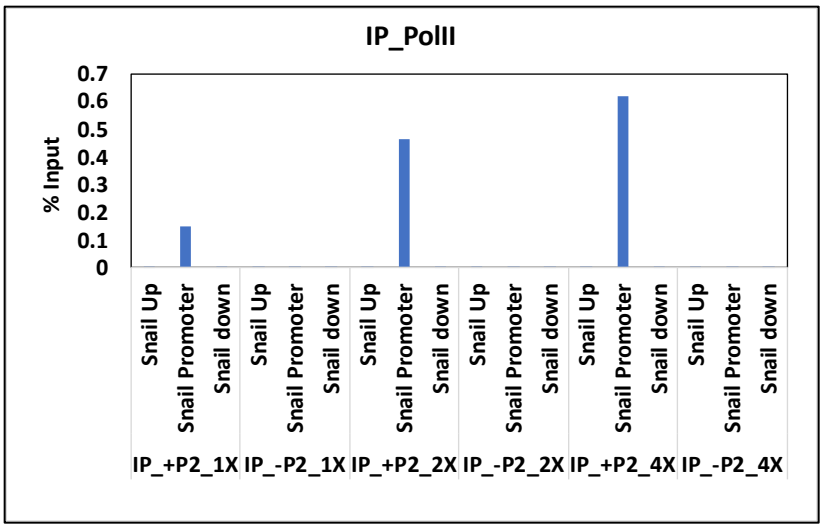


Figure 3.6. Magnetic bead standardization for ChIP-seq. ChIP-qPCR against RNA Pol II antibody showing greater pulldown of the antibody with increased concentration of bead.

The starting chromatin concentration was validated using 50µg and 100µg of MCF7 chromatin against H3K4me3 antibody (Abcam). Chromatin concentration was measured by reverse-crosslinking and extracting DNA from sonicated cells. Higher concentration of chromatin showed improved signal-to-noise ratio and 100µg chromatin was used for further ChIP experiments (Table 3.2). For transcription factor HSF1, approximately ~150µg of chromatin was used.

Table 3. 2. Table showing signal-to-noise (S/N) ratio increases by increasing starting amount of cells

	Primer	NHS (% Input)	HS (% Input)	S/N: NHS	S/N: HS
Batch1 (100ug)	PGBD2 negative control	0.01	0.03		
	PGBD2 positive control	1.74	2.26	118.94	65.98
	SH3BP5L negative control	0.06	0.10		
	SH3BP5L positive control	2.67	2.76	48.10	28.46
Batch 2 (50ug)	PGBD2 negative control	0.05	0.10		
	PGBD2 positive control	1.39	1.55	26.47	16.33
	SH3BP5L negative control	0.10	0.09		
	SH3BP5L positive control	2.69	2.19	26.10	23.60

3.6 Genome-Wide HSF1 Binding is a Better Indicator of Heat Shock Response than Gene Transcription

Previous studies show that binding of HSF1 in response to heat shock (HS) in mammalian cells occurs throughout the genome at and outside of gene promoters (Mahat *et al.,* 2016). The binding occurs only at a small proportion of potential sequences in the genome predicted through analyzing HSF1 sequence motifs. This raises a possibility that, rather than being invariant, the pattern of HSF1 binding may be flexible. To determine whether HSF1 binding is a function of the stress, we sought to find HSF1 binding in MCF7 cells treated with rapid HS (heat-dependent) and As (heat-independent). We treated MCF7 breast cancer cells with HS at 43°C for 60 minutes and performed ChIP-seq against HSF1, Pol II and promoter-associated histone mark H3K4me3 in heat-shocked and non-heat-shocked cells to investigate which of these factors are most sensitive to heat stress. We found that upon HS, MCF7 cells do not show dramatic changes in the promoter-associated histone marks or in their Pol II level (Fig 3.7A). However, HSF1 shows dramatic changes in their distribution. In contrast, massive binding of HSF1 occurs at the promoters of HS-response genes such as HSPH1 which generates Pol II wave along the gene (Fig 3.7C). We confirm, as previously reported in other mammalian cells, that there is a massive genome-wide HSF1 binding occurring both at distal regions and at promoters. (Fig 3.7B and 3.7D) Promoter histone mark H3K4me3, however, do not show any major changes with HS (Fig 3.7A).

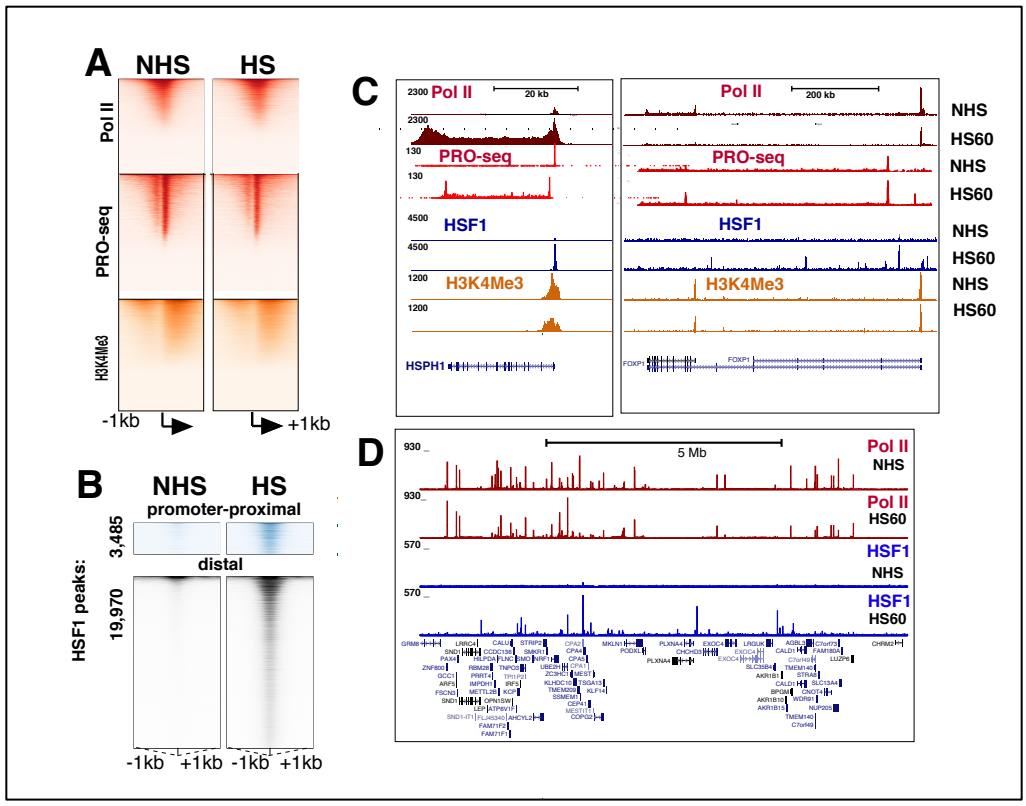


Figure 3. 7. HSF1 is more sensitive to HSR than transcription.
(A) Heatmaps showing Pol II, H3K4me3 and PRO-seq signal intensity at promoter showing no dramatic difference before or after HS.
(B) Heatmap of HSF1 peaks at promoter and distal intergenic region showing rapid binding of HSF1 to DNA with HS treatment.
(C) UCSC Browser shot of a HS-responsive gene HSPH1 showing promoter-binding of HSF1 and rapid upregulation of transcription.
(D) UCSC Browser shot of an intergenic region showing massive binding of HSF1 but no change in Pol II status or other histone marks

3.7 HS Core Responsive Genes Such as hsp70 or HSPH1 Show Decrease in H3K4me3 Peaks at their Promoters at High Temperature

Previous studies in yeast and in Drosophila have shown that hsp70 promoter loses H3K4me3 when subjected to HS[109,110,111]. However, promoter dynamics of nucleosomes in mammalian cells in response to stresses have not been looked at. ChIP-qPCR was performed against promoter and non-promoter regions of HS-responsive genes and control, non-responsive genes. Reduction of H3K4me3 mark was observed at promoters of core

HS responsive genes such as hsp70 whereas H3K4me3 at promoters of HS non-responsive genes such as ZNF692 and SH3BP5L (referred to as ZNF and SH3 respectively, in Fig 3.8 C) remained unchanged (Fig 3.8).

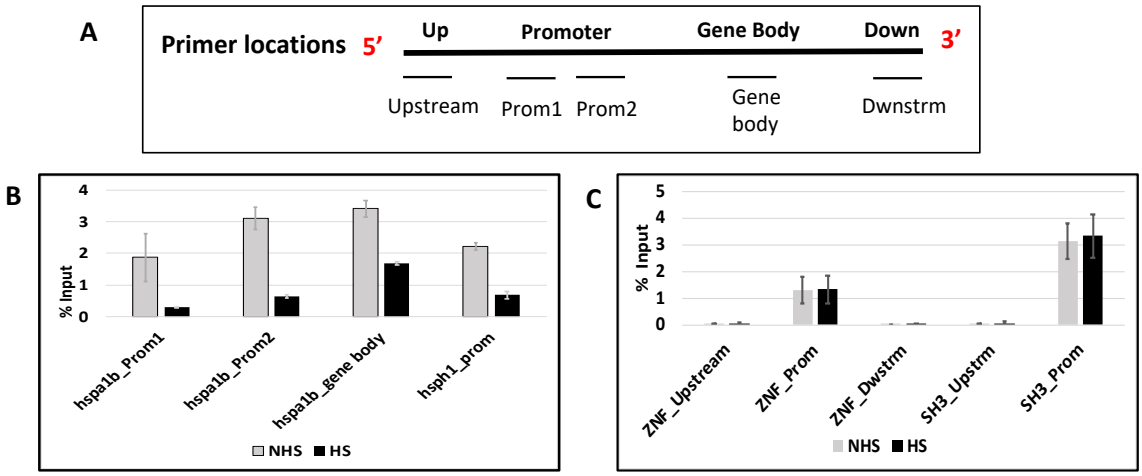


Figure 3. 8. H3K4me3 level decreases with HS only at HS-responsive genes. (A) Schematic representation of the approximate location of the primers used for ChIP-qPCR. (B) ChIP-qPCR against H3K4me3 showing decrease in the mark with HS at HS-responsive HSPA1B and HSPH1. (C) ChIP-qPCR against H3K4me3 showing no change of histone marks with HS at HS non-responsive genes

We did not observe any massive loss of total (pan) H3 at the promoters of HS-responsive gene with ChIP-qPCR against pan-H3 (Abcam) probably meaning that H3K4me3 mark is getting replaced by other marks such as H3K27ac. Interestingly, we saw gain of pan H3 mark at both upstream region and promoter regions of HS non-responsive gene. It would be interesting to see if any repressive histone modifications are added to gene-promoters of HS non-responsive genes making sure that these genes are not activated during HSR (Fig 3.9).

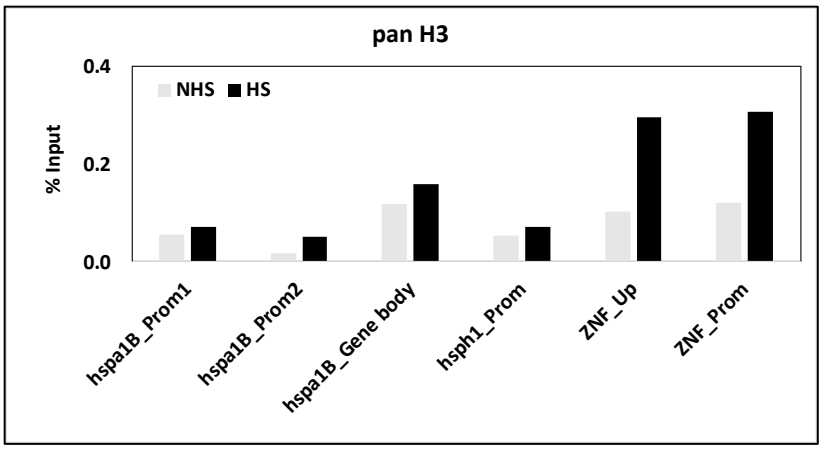


Figure 3. 9. Pan H3 level unchanged with HS. ChIP-qPCR (n=1) against pan H3 showing no major changes with HS at HS-responsive genes; however, HS non-responsive genes show an increase in pan H3.

We next asked if this decrease of H3K4me3 at promoters is linked to HSF1 binding. We divided HS-responsive upregulated genes into two groups – HSF1 dependent (363 genes) (Fig 3.10A) and HSF1-independent (374 genes) (Fig 3.10B) based on overlap of HS upregulated genes with HSF1 bound peaks at TSS +/-1kb. As expected, we found a slight dip at exactly the TSS of HSF1-dependent upregulated genes (Fig 3.10A) whereas HSF1-independent upregulated genes showed no changes at their TSS (Fig 3.10B). This is also seen in UCSC Genome Browser when we look at the promoters of core HS response genes (Fig 3.11).

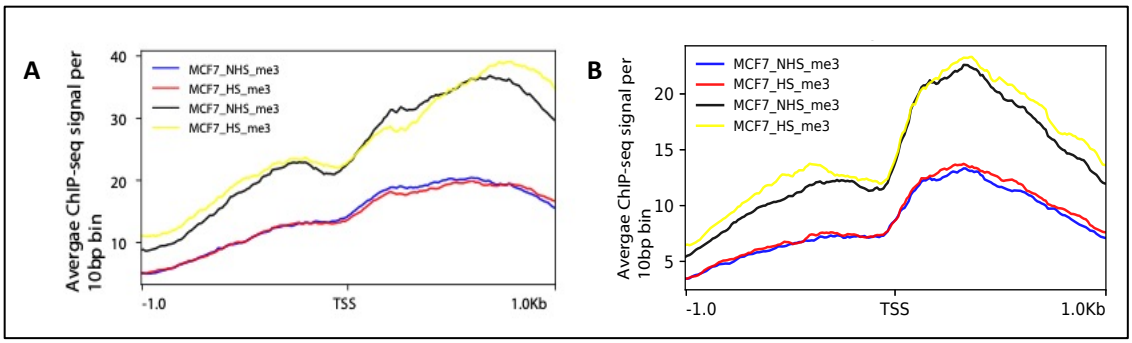


Figure 3. 10. HSF1-dependent genes show decreased H3K4me3 at promoter with HS. Profile of H3K4me3 level of HSF1-dependent (A) showing slight dip at promoter with HS and (B) HSF1-independent HSR genes showing no change of H3K4me3 level with HS. Profiles shown for both replicates.

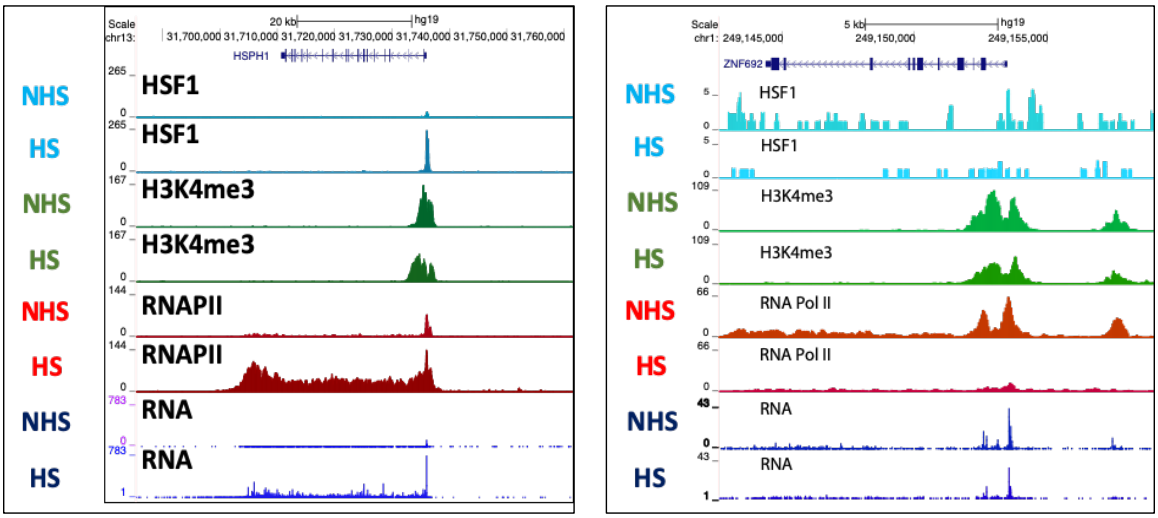


Figure 3. 11. UCSC (University of California-Santa Cruz) Browser showing profile of HSR gene HSPH1 (left) and HS non-responsive gene ZNF692 (right) with respect to HSF1, H3K4me3, RNA Pol II and PRO-seq

3.8 HS Core Responsive Genes Acquire H3K27ac Activation Mark

Transcriptionally active genes are often found to be marked by histone 3 lysine 27 acetylation (H3K27ac) mark. To investigate whether HS-responsive genes also contain H3K27ac modification at their promoters, ChIP-qPCR against HS responsive (Fig 3.12; left) and HS non-responsive genes (Fig 3.12; right) were performed. Surprisingly, we found promoters of HS-responsive genes have low H3K27ac at non heat shock (NHS) condition. Upon HS, these genes acquire H3K27ac mark whereas genes that do not respond to HS lose this mark. This is an interesting finding which might suggest that cells maintain a constant pool of histone code 'writers' and during rapid transcriptional response these writers are immediately redistributed from non-responsive genes to the activated genes.

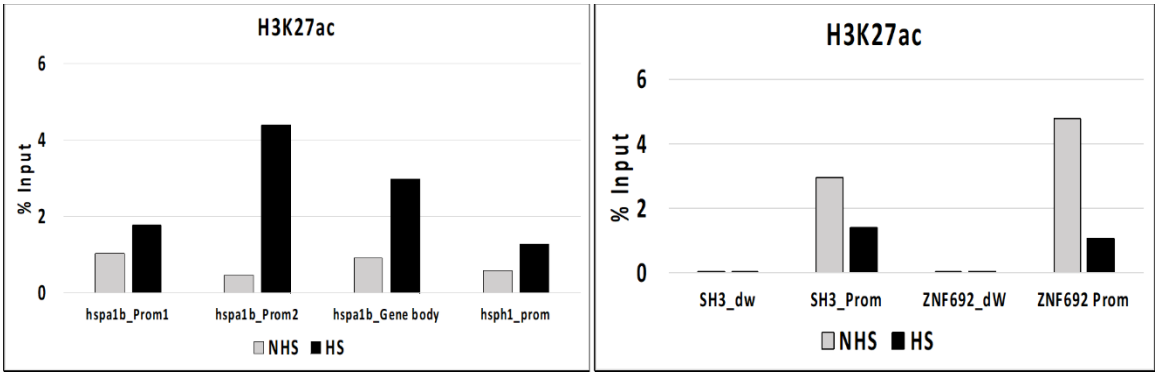


Figure 3. 12. Increase in H3K27ac enrichment at HSR genes. ChIP-qPCR against H3K27ac against HSR genes (left) and HS non-responsive genes (right) showing HSR acquire transcription-activation mark whereas HS non-responsive genes lose activation mark upon HS.

Active genes at ground state are marked by higher level of H3K4me3 than inactive genes. We categorized genes from NHS PRO-seq data into expressed and non-expressed genes based on their relative read counts at the promoter. We took genes that are present in both of our independent biological replicates at NHS and defined 'expressed genes' (which includes both active and poised genes) as genes with absolute value for promoter counts more than 10 and 'non-expressed genes' as genes with absolute value for promoter counts less than 10. We then measured H3K4me3 level at these genes' promoters and found that expressed genes have a much higher level of H3K4me3 at basal condition than the non-expressed ones. This observation supports the notion that H3K4me3 modification marks only transcriptionally active or poised genes (Fig 3.13). This result also shows that majority of the already active genes at ground state do not exhibit a proportional increase or decrease in H3K4me3 mark with HSR, indicating that H3K4me3 marks are stable and do not undergo rapid changes (Fig 3.13).

3.9 MCF7 Cells Elicit Robust HSR when Exposed to Arsenic (As)

Arsenic is a metalloid that is found in soil, air and water in low amount. Chronic exposure to arsenic can lead to cancer whereas arsenic has also been used as drug in treating acute promyelocytic leukemia (AML). Sodium arsenite, a trivalent inorganic form of Arsenic, causes HSR in mammalian cells and triggers HSF1 pathway.

Similar to HS, MCF7 cells were treated with 500 μM sodium meta-arsenite (a compound of arsenic) (Sigma S7400) for 0 minutes, 15 minutes, 40 minutes, 60 minutes, and 180 minutes, and changes in gene expression was measured with RT-qPCR against the same genes as those used for validating HS (Fig 3.3). Response to Arsenic was further

validated by performing a dose-curve using 0 μM, 100 μM, 200 μM and 500 μM of Arsenic for 60 minutes and measuring gene-expression with RT-qPCR (Fig 3.14).

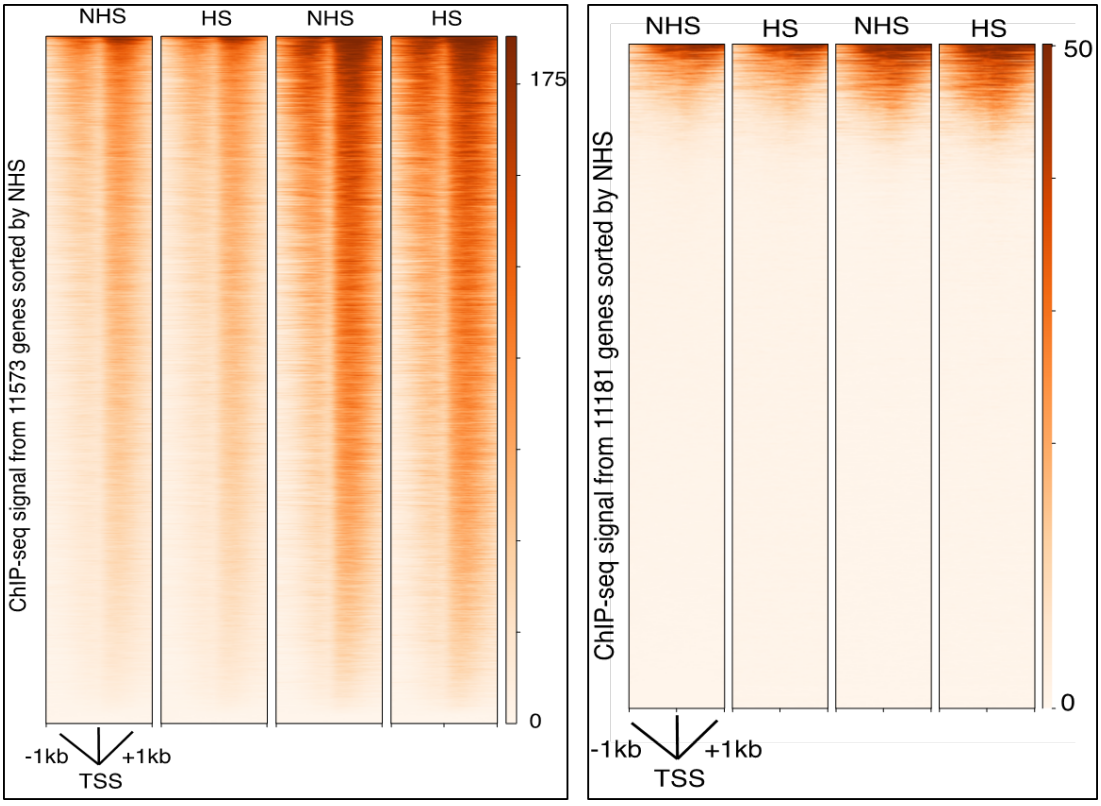


Figure 3. 13. H3K4me3 is a marker of active genes at ground state in MCF7 cells. Heatmap against H3K4me3 showing there is much less binding of H3K4me3 at inactive genes (left) than active genes (right) at ground/basal state; although this H3K4me3 state remains mostly unaffected genome-wide after HS.

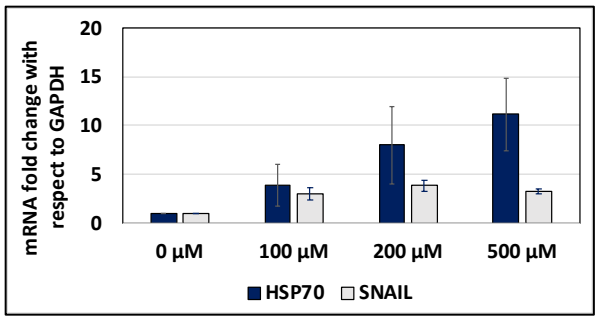


Figure 3. 14. Dose-response curve of As in MCF7 Cells. RT-qPCR showing increased dose of As elicits a robust HSR within 60 min.

3.10 Genome-Wide HSF1 Binding Follows an HSR Program Independent of Temperature

We wondered if HSR induced by a temperature-independent stimulus would follow the same pattern as temperature-induced HSR. We exposed MCF7 cells to 500 µM of sodium arsenite for 60 minutes which is the exact time we exposed cells to heat stress. We validated that the response of MCF7 to arsenic (As) is similar to heat-shock by RT-qPCR in terms of fold activation and timing of response (Fig 3.15A). Following validation of HS and As treatment conditions, we then performed ChIP-seq against HSF1 and Pol II in Arsenic treated MCF7 cells, and compared HSR between these heat-dependent and heat-independent stresses (Fig 3.15B). HSF1 showed robust binding to DNA both in HS and As with ~ 95% of peaks newly appearing in either treatment condition compared to NHS condition. As, similar to HS, induces HSF1 binding to promoters of HS-responsive genes and activates their expression. We found that a majority (~ 61%) of HSF1 peaks induced in response to HS or As are common between these two treatments (Fig 3.15C). We then plotted heatmap taking HSF1 peaks common to both HS and As, and HSF1 peaks appearing either in HS or As condition (Fig 3.15D). Surprisingly we found that peaks called exclusively in HS are actually present in As and vice-versa indicating that some of the

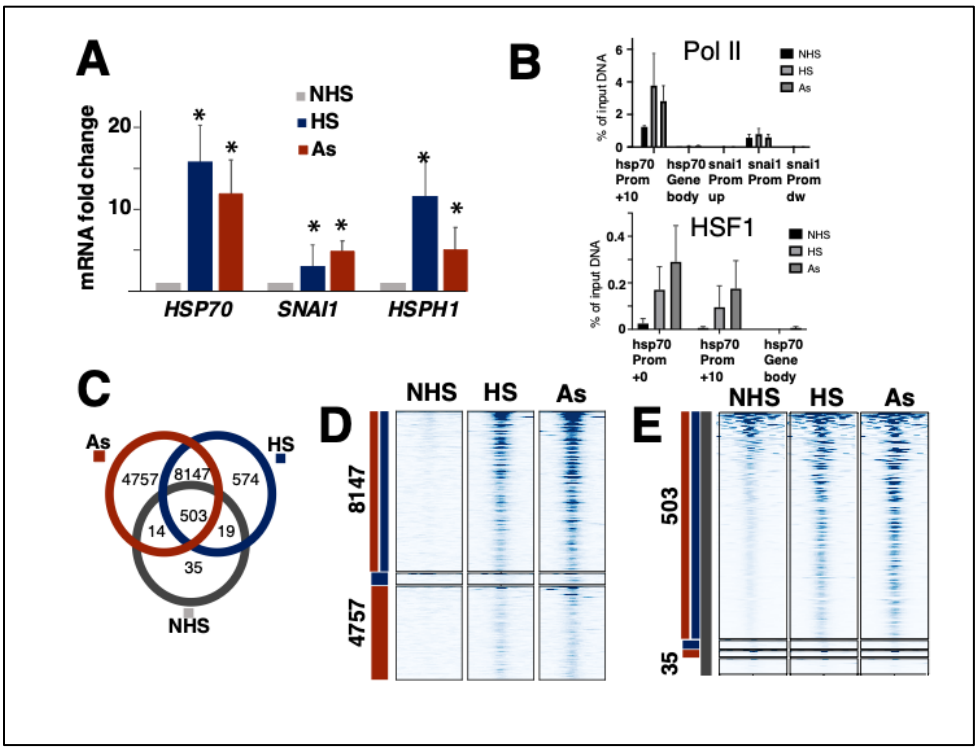


Figure 3. 15. MCF7 cells follow an HSR program that is independent of heat.
A. RT-qPCR showing MCF7 elicits HSR to heat-independent stress As, similar to HS.
B. ChIP-qPCR against Pol II and HSF1 showing enrichment of both at HS-responsive gene promoter and negative control regions show no enrichment.
C. Overlap of HSF1 bound peaks at NHS, HS and As state showing majority of the peaks overlap within the conditions.
D. Heatmap of the HSF1-bound peaks showing peaks called exclusively in HS are also present in As, and vice versa indication the similarity of the HSR program between the two treatments.

unique peaks called might be artefacts of the peak calling software rather than indicating biological differences. Furthermore, we found that 88% of the peaks called at NHS condition are present in both HS and As indicating HSF1 binding regions might be predetermined in a cell type and remain invariant in heat-dependent or heat-independent stresses; with exposure to stresses it is the same program that gets amplified.

To further validate similarity between HS and As induced HSR, we plotted Spearman correlation among NHS, HS and As Pol II (Fig 3.16; left) and HSF1 (Fig 3.16; right) samples, respectively and as expected, we found high correlation among the treatments and the replicates.

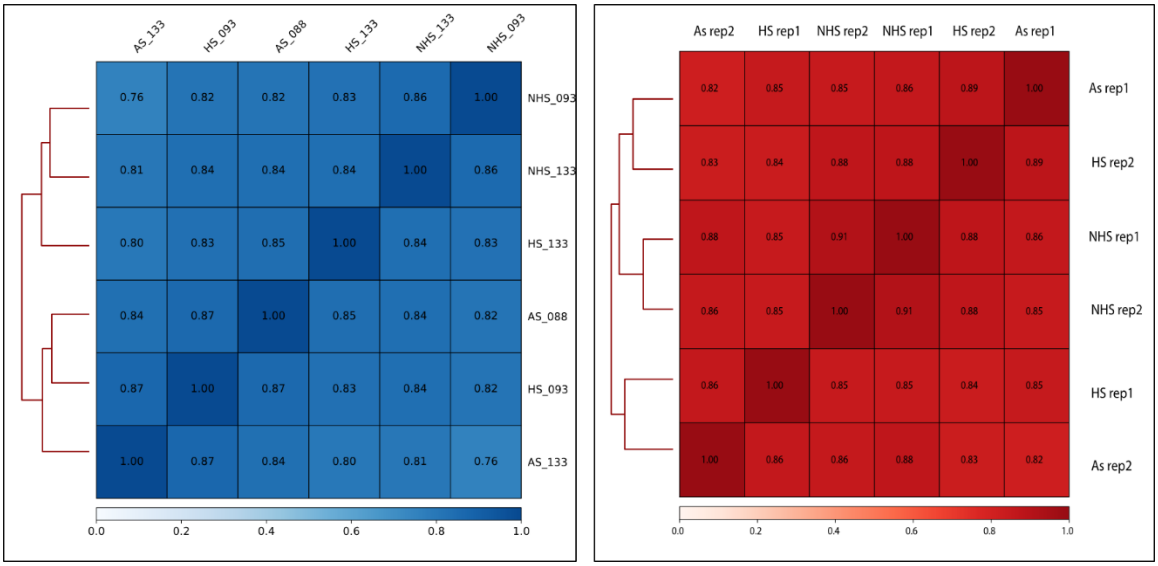


Figure 3. 16. HSF1 and Pol II ChIP-seq samples are highly correlated among different conditions in MCF7 cells. Spearman correlation of HSF1 (left) and Pol II (right) indicated in the squares show similarity in HSR program between different conditions

3.11 Genome-Wide HSF1 Binding Follows Cell-Type Specific Programs

To understand if HSR follows the same program even across cell-type as it does for heat-dependent and heat-independent stresses, we compared HSR in MCF7 breast cancer cells to K562 chronic myeloid leukemia cells. K562 cells have been previously used for studying HSR and is known to induce robust HSR upon heat stress. We found that both cell lines exposed to HS and As induce phosphorylation of HSF1 (Fig 3.17A). We performed ChIP-seq against HSF1 and Pol II in HS and As-treated K562 cells. Comparison of HSF1 peaks between MCF7 and K562 in these two conditions clearly shows there is a

difference in genome-wide HSF1 distribution between the two cell types exposed to the same stresses (Fig 3.17 A). However, both cell types induce similar HSR programs in either HS or As. Principal component analysis of the two cell types under NHS, HS and As induction points out cell type specific clustering rather than treatment-specific clustering (Fig 3.17B). Correlation analyses of HSF1 signal in peaks between cell types strongly support the cell-specificity of HSF1 distribution; when HSF1 peaks between HS and As are compared in any of the cell type, Spearman R is more than 50% in both cases. However, when we compare MCF7 versus K562 for the same treatment (either HS or As) Spearman R is below 40% (Fig 3.17 C&D).

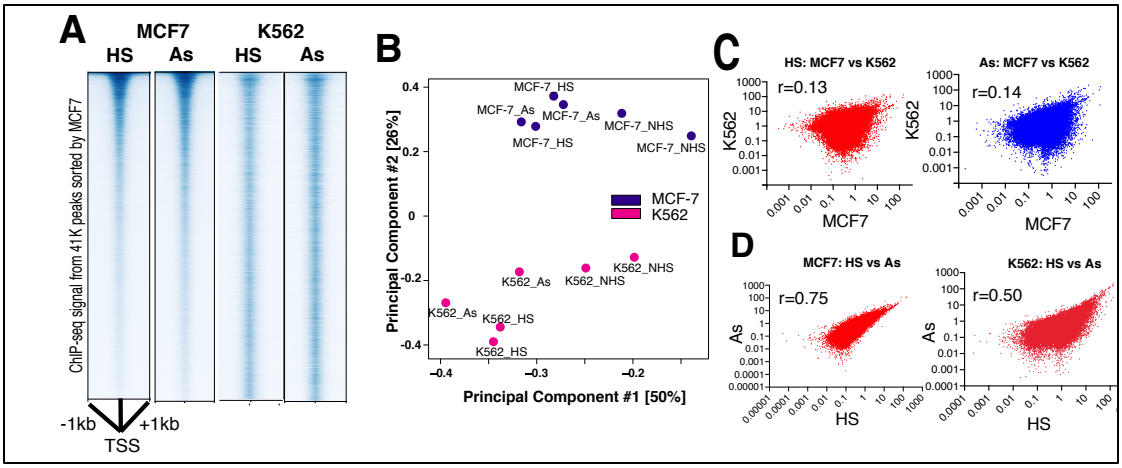


Figure 3.17. HSR is cell specific.
(A) showing distinct HSF1 enriched regions in MCF7 HS or As treated versus K562 HS- or As-treated cells.
(B) Principal component analysis (PCA) plots showing cell types cluster together irrespective of the treatment for HSF1 ChIP-seq.
(C & D) Spearman correlation plotted in log10 scale showing difference between cell-type for a particular treatment (C) is greater than difference between two treatments in the same cell type (D).

3.12 ChIP-seq Results Against HSF1 is not an Artefact of the Antibody Used

Mendillo et al., 2012 had reported that in several cancer cell lines HSF1 follows a cancer-specific heat-shock program which occurs at basal state without HS induction and is different from heat-dependent heat-shock program. They generated ChIP-seq data against polyclonal HSF1 antibody from SantaCruz in several cell types and primary cells including MCF7 at NHS condition. However, the antibody has been discontinued by the manufacturer. We first titrated the newly available monoclonal HSF1 antibody from SantaCruz with polyclonal HSF1 from Enzo with MCF7 cells treated with Arsenic. We validated the result with ChIP-qPCR with both antibodies. SantaCruz antibody yielded poor pulldown of chromatin but with Enzo successful pulldown of HSF1 was achieved (Fig 3.18). Enzo antibody was used for further HSF1 ChIP experiments.

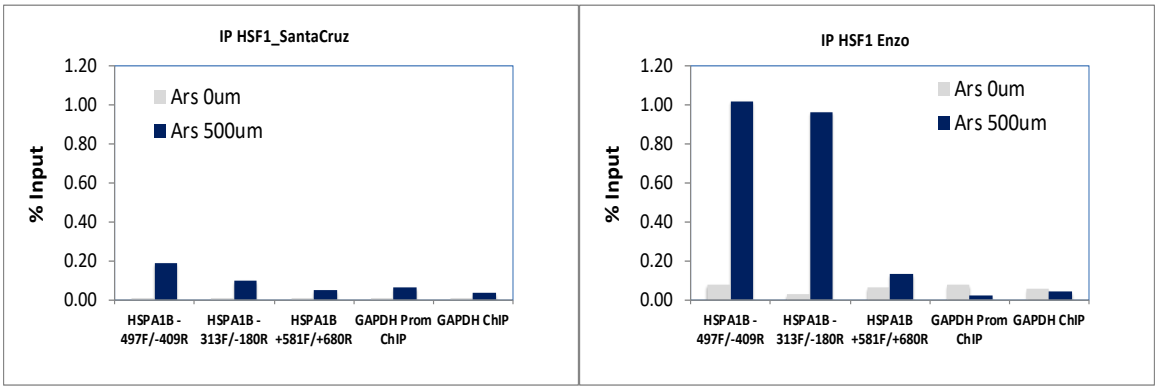


Figure 3.18. Standardization of HSF1 antibody: monoclonal versus polyclonal. ChIP qPCR against HSF1 in MCF7 As treated cells using monoclonal SantaCruz antibody (left) resulted in lesser enrichment than polyclonal HSF1 Enzo antibody (right).

To confirm that pull-down with HSF1 antibodies from SantaCruz versus Enzo is not a technical consequence of the ChIP-seq protocol, we performed ChIP qPCR with Pol II control antibody which we had already validated before (Fig 3.19).

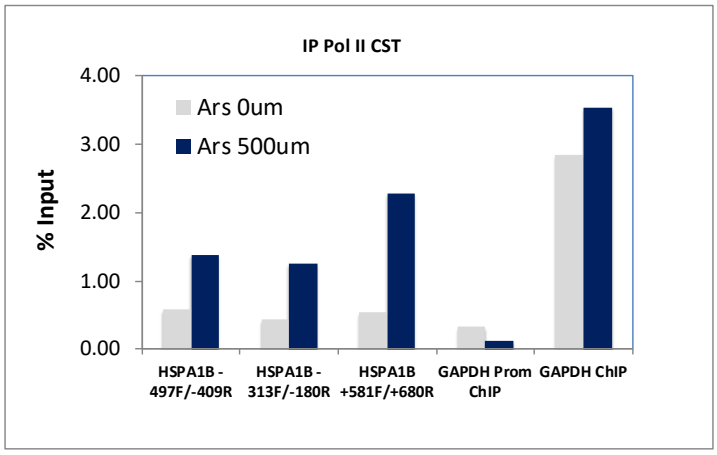


Figure 3.19. Validation of ChIP protocol during HSF1 antibody test: ChIP-qPCR against previously validated Pol II antibody showed enrichment of Pol II with As treatment thus validating the ChIP protocol

To confirm that DNA binding by HSF1 is not an artefact of the specific antibody, published HSF1 ChIP seq data were downloaded from GEO database (Table 3.3) and analyzed using the same bioinformatic pipeline as our data.

Table 3. 3 Downloaded ChIP-seq data from NCBI database with their accession number, sample IDs and references.

GEO Accession Number	Sample	Publication
GSM951882	MCF7_HSF1_ChIPSeq_1	Mendillo *et al*, *Cell* 2012 Aug 3;150(3):549-62
GSM951883	MCF7_HSF1_ChIPSeq_2	Mendillo *et al*, *Cell* 2012 Aug 3;150(3):549-62
GSM1065711	HSF1_non-treated cycling K562	Vihervaara A *et al*. *Proc Natl Acad Sci U S A* 2013 Sep 3;110(36):E3388-97
GSM1065714	HSF1_heat-treated cycling K562	Vihervaara A *et al*. *Proc Natl Acad Sci U S A* 2013 Sep 3;110(36):E3388-97

Spearman R gives a correlation of more than 70% between HSF1 peaks found in our ChIP-seq data and in MCF7 NHS samples generated by Mendillo et al., when both data are processed by us using the same parameters. (Fig 3.20; left). These peaks are similar to those noted by Mendillo et al. (Fig 3.20; right).

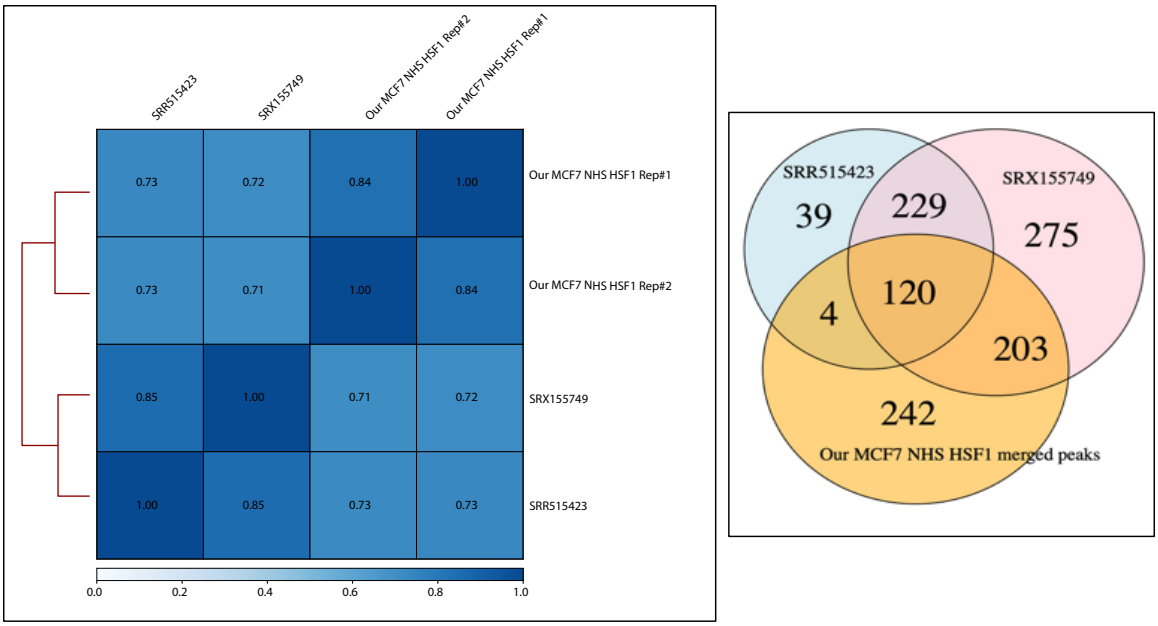


Figure 3.20. MCF7 NHS HSF1 ChIP-seq data have high correlation with published ChIP-seq data. Heatmap with spearman correlation values in small squares between our NHS HSF1 ChIP-seq and published NHS HSF1 ChIP-seq in MCF7 cells showing high correlation among the samples (left). Overlap of peaks called in published HSF1 ChIP-seq data and our ChIP-seq data show more than 50% overlap of called peaks.

We have unpublished ChIP-seq data against old SantaCruz HSF1 antibody which we re-analyzed using the same pipeline and showed a correlation of ~ 50% with our data with Enzo antibody (Fig 3.21 right) and majority of HSF1 peaks overlapped in NHS and HS condition of MCF7 cells (Fig 3.21 left).

We further plotted heatmaps using peaks that were called in both datasets and confirmed that peak location is similar with both antibodies (Fig 3.22) in the same cell type MCF7.

For validation of K562 HSF1 ChIP-seq data, we downloaded previously published data (Vihervaara et al., 2013) (Table 3.3) in K562 cell line and ran it through the same

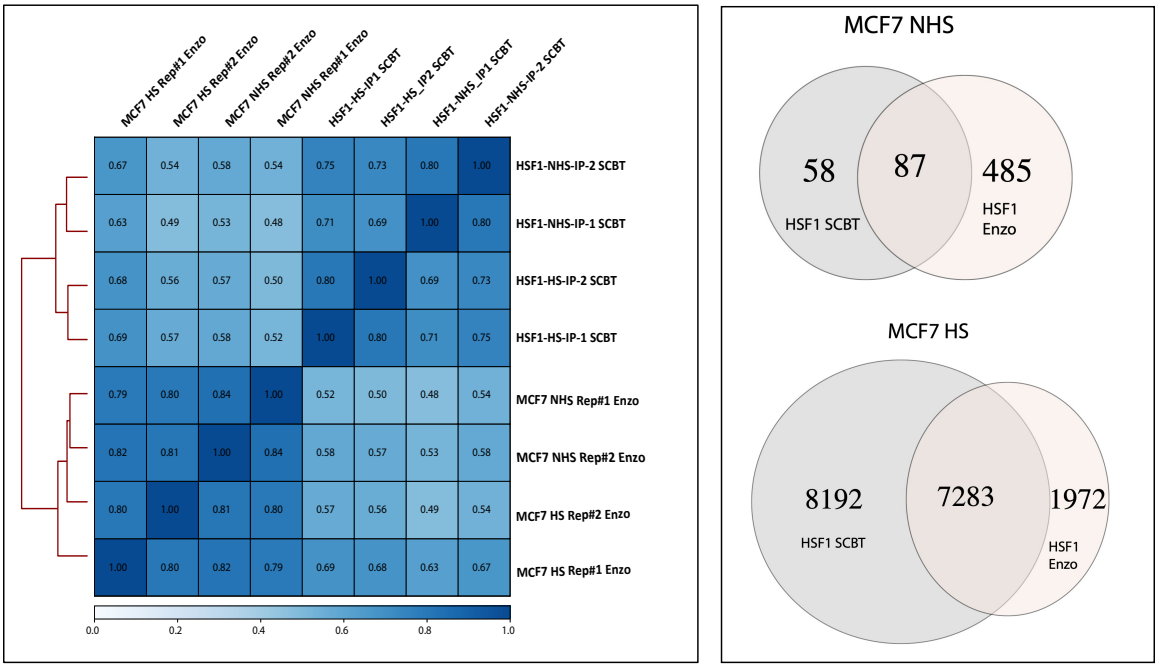


Figure 3.21. New polyclonal HSF1 Enzo antibody shows good correlation with old polyclonal HSF1 SantaCruz antibody. Heatmap with spearman correlation values in small squares between old and new HSF1 antibodies in MCF7 NHS and HS samples (left). Venn diagrams showing 60% of NHS peaks and 78% of HS peaks overlap with each other (right)

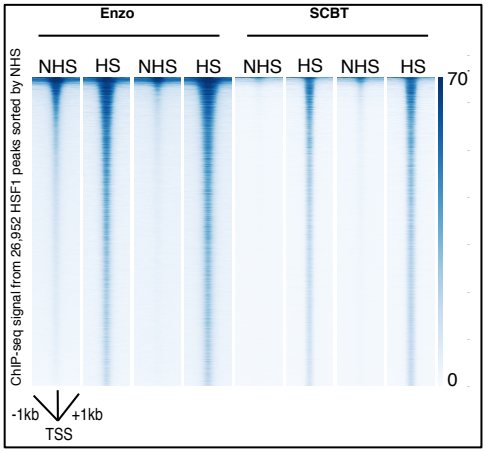


Figure 3.22. HSF1 enrichment at similar genomic regions detected by both SCBT and Enzo antibodies. Heatmap showing HSF1 ChIP-seq signal at NHS and HS against Enzo and SCBT HSF1 showing similar genome-wide distribution.

pipeline we used for our K562 HSF1 ChIP-seq analysis. Approximately 50% correlation was found with Spearman statistical analysis between published data and our samples with majority of the HSF1 peaks common between both samples (Fig 3.23)

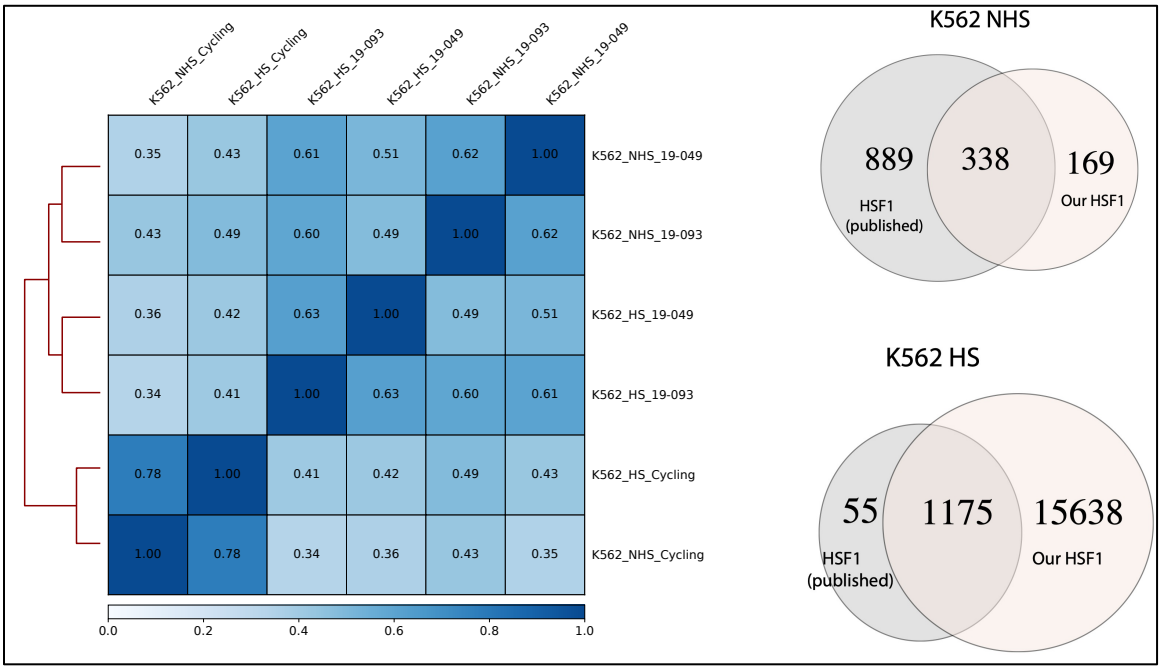


Figure 3.23. HSF1 enrichment between published K562 ChIP-seq data and our K562 ChIP-seq data is well correlated. Heatmap with spearman correlation values in small squares between published and our HSF1 ChIP-seq in K562 NHS and HS samples (left) show ~50% correlation. Venn diagrams showing 66% of NHS peaks and 95% of HS peaks overlap with each other (right).

3.13 HSF1 is Enriched at Promoters in Both Treatments in Both Cell Lines

As defined using UCSC hg19 database, promoters comprise ~2.35% of the total human genome. Though HSF1 binds extensively throughout the genome, in both MCF7 and K562, HSF1 binding nevertheless shows enrichment at promoters over non-promoter regions. In heat-shock, 16% and 10% of total HSF1 peaks are bound to promoters in MCF7 and K562, respectively. Similarly, in Arsenic, 13% and 10% HSF1 peaks are bound at

promoters in MCF7 and K562 respectively. These observations are consistent with previous studies in other cell types (Mahat et al., 2016; Vihervaara et al., 2013) demonstrating preference of HSF1 for promoters. Majority of the remaining HSF1 peaks are found at intergenic regions, suggesting that HSF1 might bind strongly to enhancers and can regulate gene expression from promoter-distal regions (Fig 3.24)

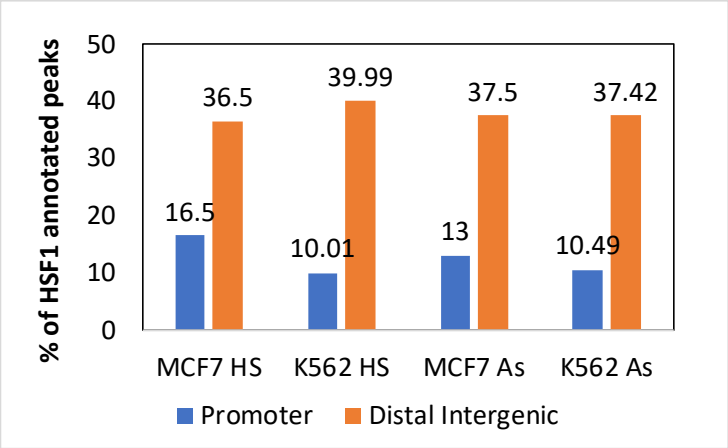


Figure 3.24. Binding affinity of HSF1 to different genomic regions. HSF1 has a strong binding affinity towards gene promoters, although majority of the HSF1 peaks appear at intergenic regions.

3.14 Differences in HSF1 Peaks between MCF7 and K562 Arise from Distal Intergenic Regions of the Human Genome

The differences in location and intensity of HSF1 peaks led us to ask where the differences are coming from. As we have previously shown, binding of HSF1 in response to both treatments occurs not only at promoters but throughout the genome. We therefore annotated promoter-proximal (within 1kb of TSS) and promoter-distal peaks (outside TSS +/-3kb) and found both for HS and As treatments, peaks that are closer to promoters overlap more between cell types versus peaks that are distal to promoters (Fig 3.25).

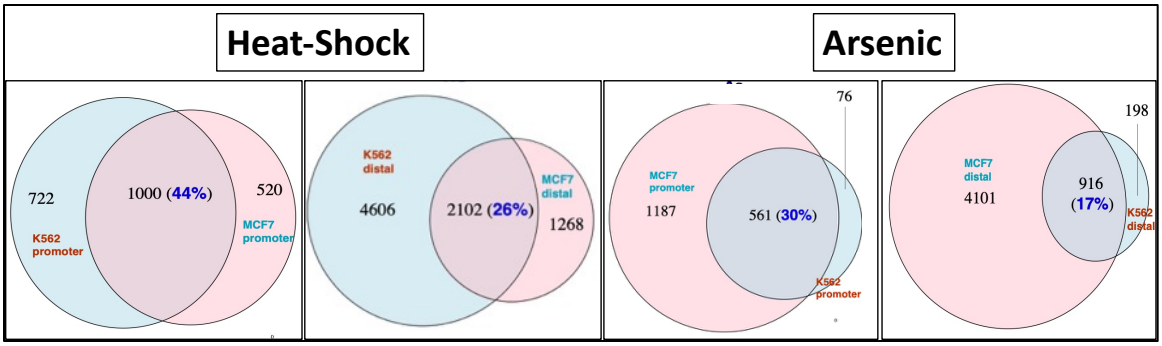


Figure 3.25. Variations in HSF1 distribution between cell type arise from promoter-distal regions. Venn diagrams showing overlap of promoter and distal region in HS (left) and As (right) between MCF7 and K562 reveal differences in intergenic HSF1 binding are more prominent than differences in HSF1 binding to gene promoters.

To detect correlation between HSF1 peaks that were present at NHS versus HSF1 peaks present at HS or As in both cell lines, we calculated Spearman correlation based on the signal generated from bins pertaining to all called peaks present in all replicates and all conditions using deepTools multibigWigsummary and GraphPad PRISM software. We found that in both treatments, MCF7 has a higher correlation between NHS and HS peaks than K562 (Fig 3.26). These results probably indicate that with stress, K562 gains new HSF1 peaks whereas in MCF7, HSR is only amplification of signal already present at ground state.

We further dissected these correlation plots into changed and unchanged HSF1 peaks by taking a cut-off value derived from the average signal from peaks that appear at the tip of 'narwhal's tusk'. Peaks that fell below this cut-off were designated as unchanged HSF1 peaks whereas peaks that were above the cut-off were designated as changed HSF1 peaks. Principal component analyses of these changed (Fig 3.27A) and unchanged HSF1 peaks (Fig 3.27B) show that peaks that change or do not change with stress cluster together indicating all HSF1 peaks are cell-type specific.

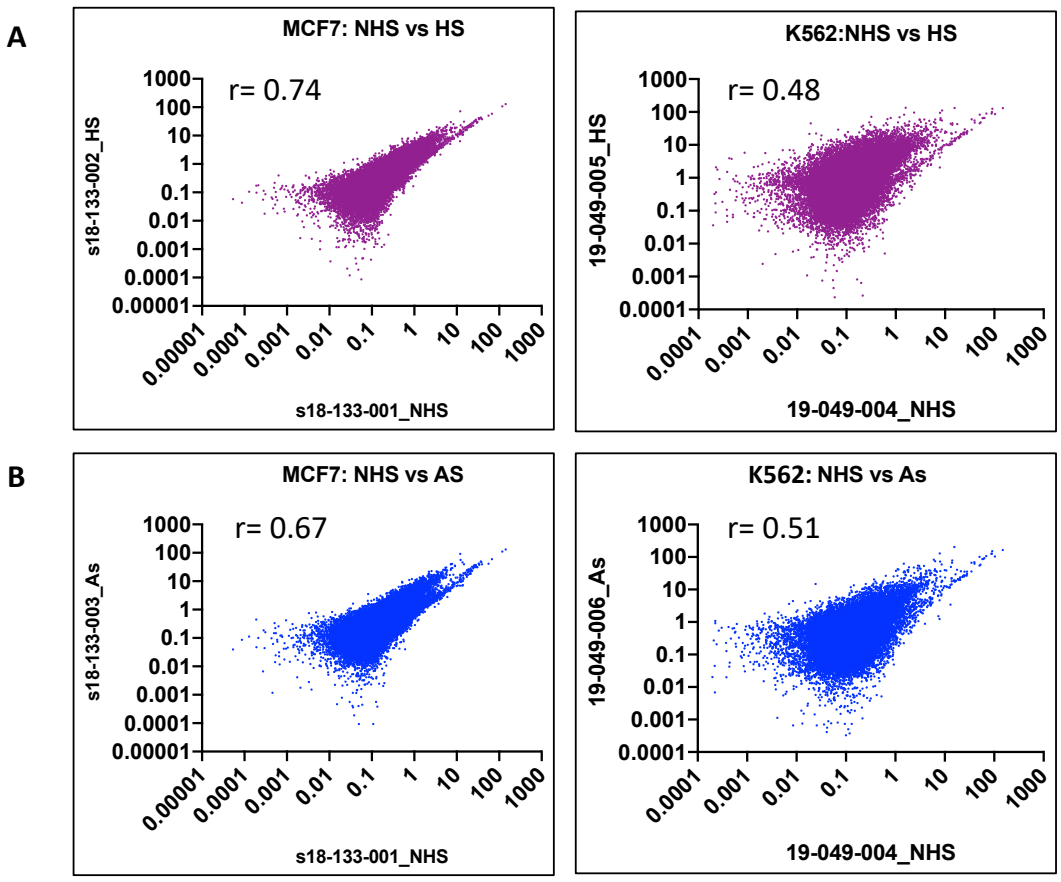


Figure 3.26. MCF7 HSF1 peaks gained as a result of HSR have a higher correlation to ground state HSF1 peaks than K562. (A) Scatter plot distribution of NHS vs HS HSF1 peaks in MCF7 and K562 with their Spearman correlation co-efficient. (B) Scatter plot distribution of NHS vs As HSF1 peaks in MCF7 and K562 with their Spearman correlation co-efficient.

To find if HSF1 is binding to different motifs in these two cell types, we did HOMER motif analyses on NHS, HS and As conditions of MCF7 and K562 cells and found, as observed before, HSF1 binding is very specific and binds to its consensus motif in all these conditions (Fig 3.28). Hence, the HSF1-binding motif or the DNA sequence is not the primary determinant of HSF1 binding specificity.

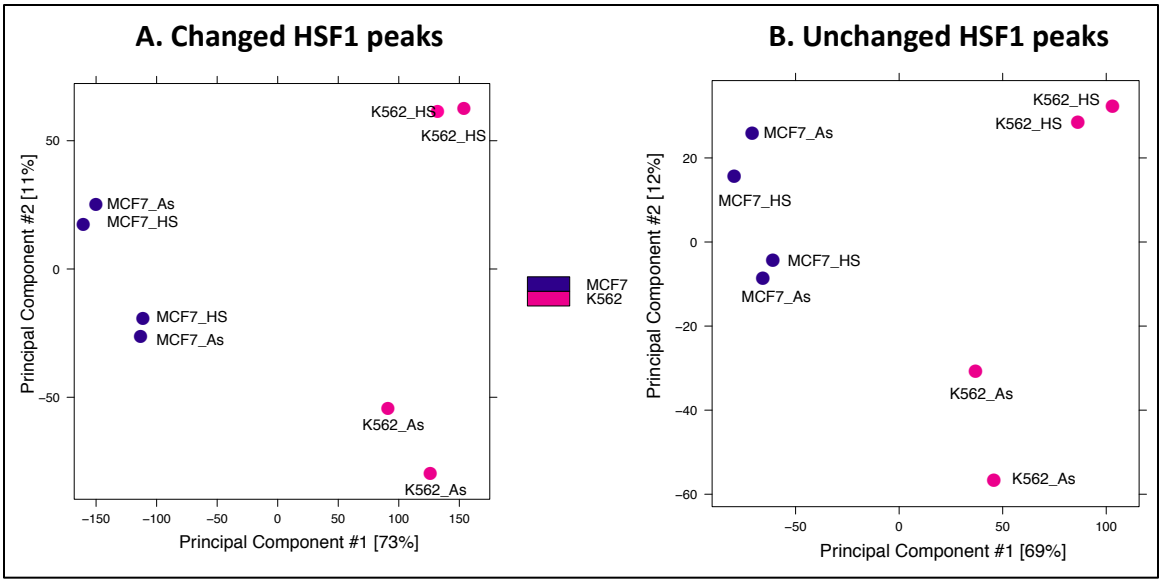


Figure 3.27. HS non-responsive HSF1 peaks are more specific to cell type than HS responsive HSF1 peaks. PCA plot showing both changed (A) and unchanged (B) HSF1 peaks show cell-type specific clustering.

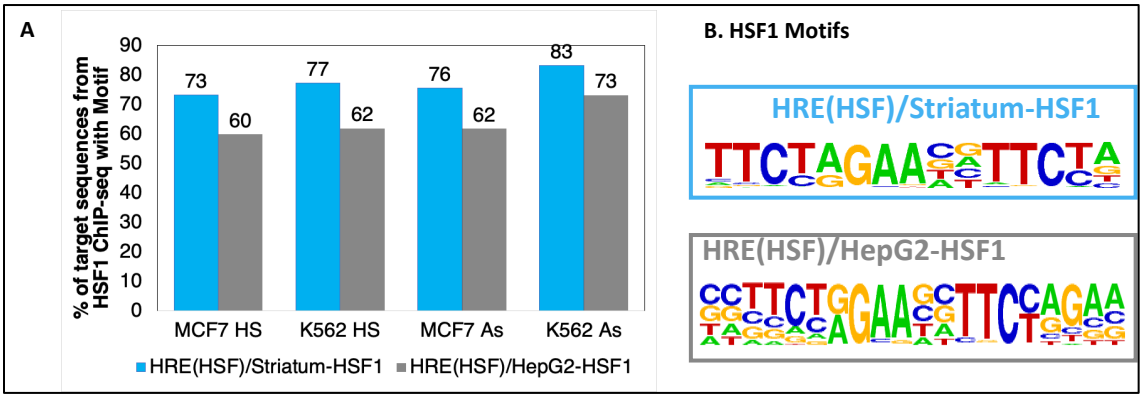


Figure 3.28. HSF1 binding is highly specific. Percentages of HSF1 motif occurrences in MCF7 and K562 cells under different stresses (A). HSF1 motif sequences found in our ChIP-seq data (B)

3.15 Chromatin Openness is not the Major Determining Factor for HSF1 Binding

We next asked how does HSF1 specifically bind to only 1% of its target sites and whether its binding sites are different among different cell types. We next hypothesized that the differences in HSF1 binding in these two cell types might be due to difference in chromatin openness between the two cell lines in response to stresses. To test this hypothesis, we performed ATAC-seq on HS and As treated MCF7 and K562 cells. Spearman correlation was greater than 50% among different conditions in the same cell type (Fig 3.29).

Surprisingly we found that only minority of the ATAC seq peaks overlap with HSF1-bound peaks in both cell lines under both treatments (Fig 3.30).

We further isolated peaks that occur at promoter (+/- 1kb) and overlapped HSF1 and ATAC peaks for MCF7 cells but only a minority of peaks were at the same location between each other (Fig 3.31).

We wondered if we were unable to determine overlap between HSF1 and ATAC peaks because of not selecting the correct distance ("maxgap" parameter in findOverlapsofPeaks package) for overlap. To negate that possibility, we randomly selected a sample (K562 As ATAC versus K562 As HSF1) and mapped overlap over a range of distance from peaks that are 100 bp apart to peaks that are 100,000 bp apart. Interestingly, increasing the range of overlap did not result in substantial increase in overlapping peaks, indicating that ATAC signal and HSF1 binding are probably independent events in the cell (Fig 3.32).

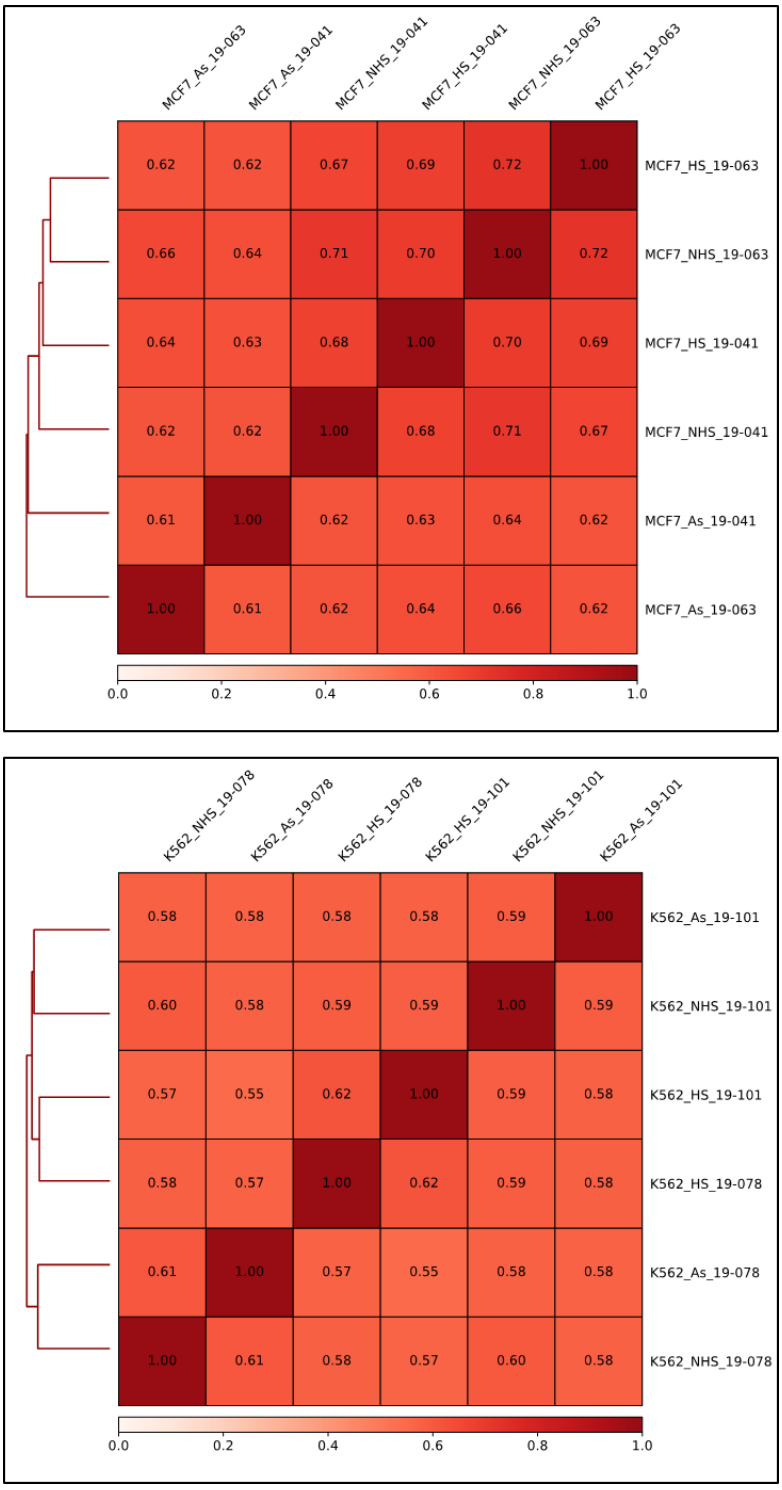


Figure 3.29. ATAC-seq libraries in different treatment conditions show high correlation in same cell-type. Heatmap with spearman correlation values in small squares between NHS, HS and As in MCF7 (top) and K562 (bottom)

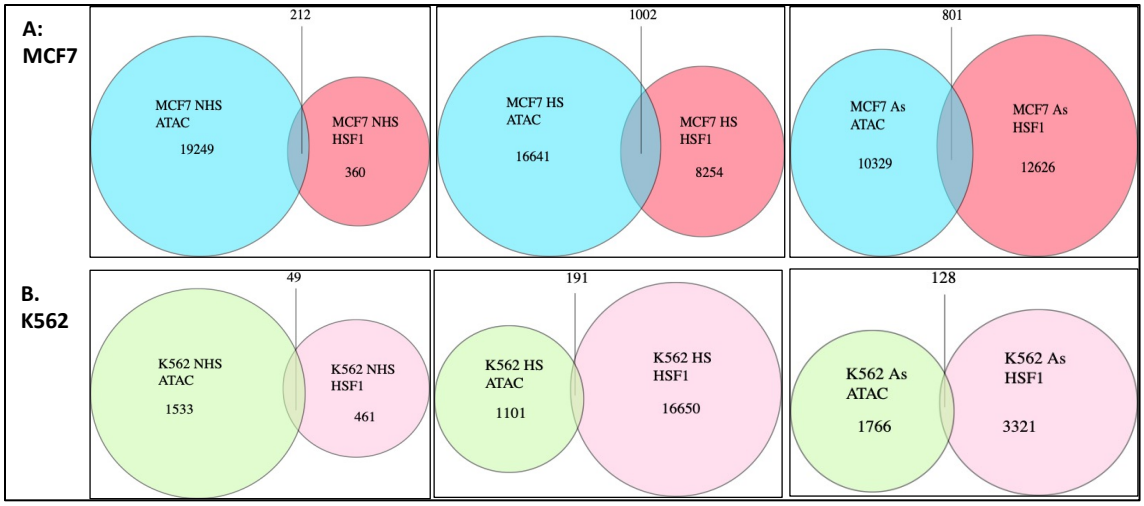


Figure 3.30. HSF1 peaks do not overlap with ATAC-seq peaks. Venn diagram between HSF1 peaks and ATAC-seq peaks in MCF7 (A) and K562 (B) showing majority of the peaks are unique between HSF1 and ATAC.

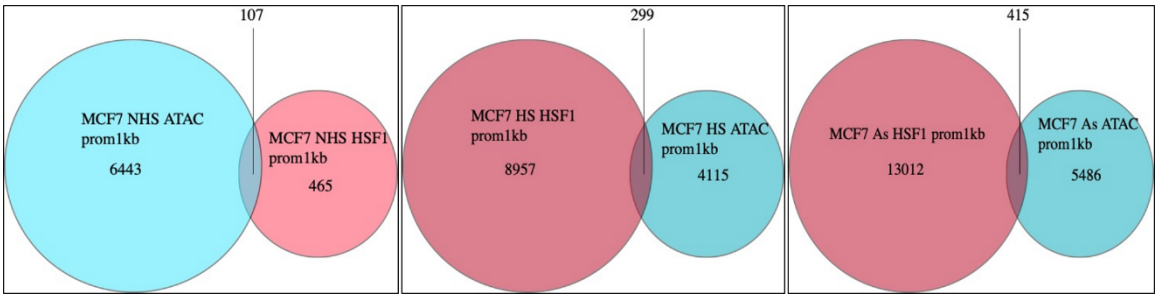


Figure 3.31. DNA accessibility is not a major determining factor for HSF1 binding. peak overlap between HSF1 peaks do not overlap with ATAC-seq peaks. Venn diagram between HSF1 peaks and ATAC-seq peaks in MCF7 in different treatment conditions showing even promoter peaks do not overlap significantly with each other.

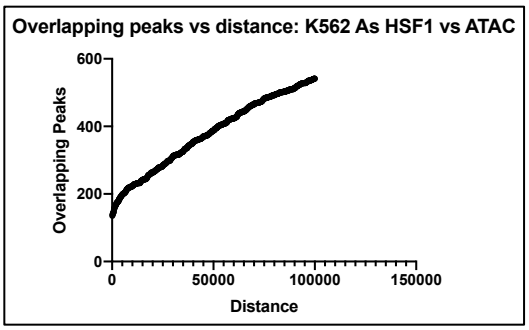


Figure 3.32. Decreasing overlapping distance criteria (maxgap) do not increase HSF1 and ATAC-peak overlap. Graph showing no relation between ATAC and HSF1 peaks, even when maximum gap between peaks is increased to 1,50,000.

To further validate this observation, published data (Vihervaara *et al*., 2017) on predicted genome-wide HSF1 binding sites in K562 NHS and HS conditions were downloaded, analyzed, and overlapped with our ATAC seq data and HSF1 ChIP-seq data from K562 NHS and HS. There is an overlap of 95% between our ChIP-seq data and their predicted HSF1 binding data (Fig 3.33; left); and ATAC-seq peaks did not overlap with published HSF1 predicted regions (Fig 3.33; right), confirming that HSF1 does not preferentially bind to open regions.

3.16 Chromatin Openness at Intergenic Regions is Distinct in Different Cell Types

We next asked if open chromatin regions detected by ATAC are similar between the two cell types. We found that MCF7 cells have much more ATAC peaks than K562. However, we found at ground state, 72% of ATAC peaks in K562 overlap with that of MCF7. When cells are exposed to HS or As, this overlap decreases in both cases to 51% and 63% respectively, indicating cell-type specific response to both stresses (Fig 3.34).

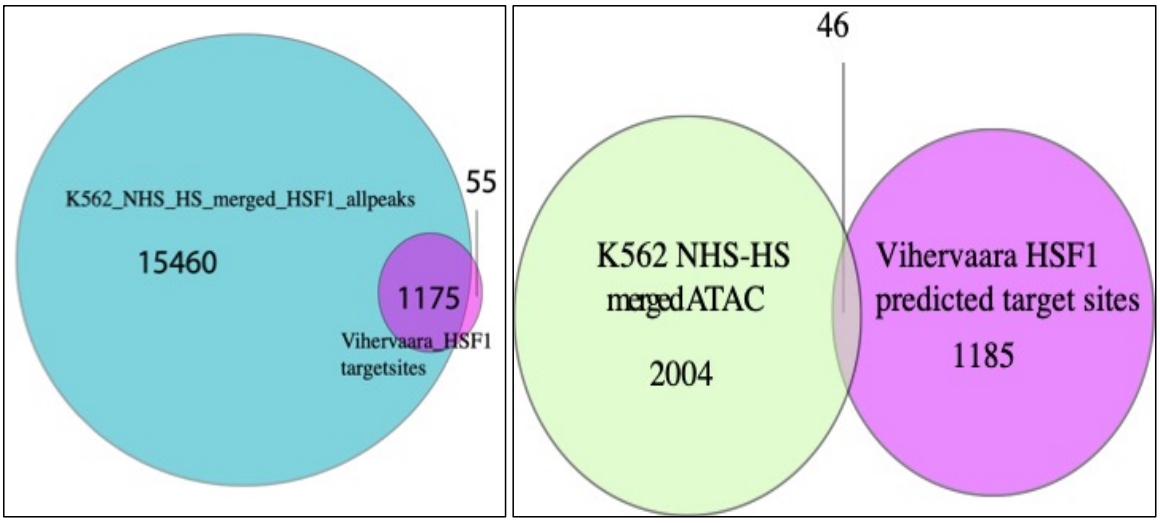


Figure 3.33. ATAC-seq peaks do not overlap with published K562 HSF1 binding sites (predicted). Venn diagram between HSF1 peaks from NHS and HS samples showing 90% overlap with published predicted HSF1 binding sites in K562 (left); Venn diagram between ATAC peaks from NHS and HS samples showing barely any overlap with published predicted HSF1 binding sites in K562 (right).

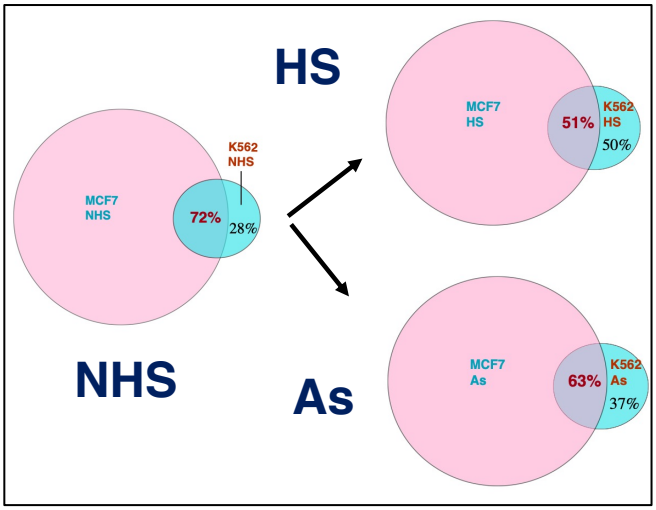


Figure 3.34. Overlap in chromatin open regions between two cell types decrease with stress. ATAC-seq showing 72% of open chromatin regions in K562 are similar to that of MCF7. However, with HS and As there is local chromatin changes resulting in decrease of overlapping regions between the cells

We wondered about the genomic locations of these overlapped peaks versus the non-overlapped ATAC peaks. These overlapping and unique peaks for each cell line and each condition were annotated and as expected, more than 80% of the overlapped peaks came from promoter regions for both stresses whereas unique peaks were more frequently found at the distal intergenic regions. These results indicate that potential enhancers in the distal intergenic regions might be more cell-type specific and might contribute to determine a cell's identity rather than the promoters (Fig 3.35).

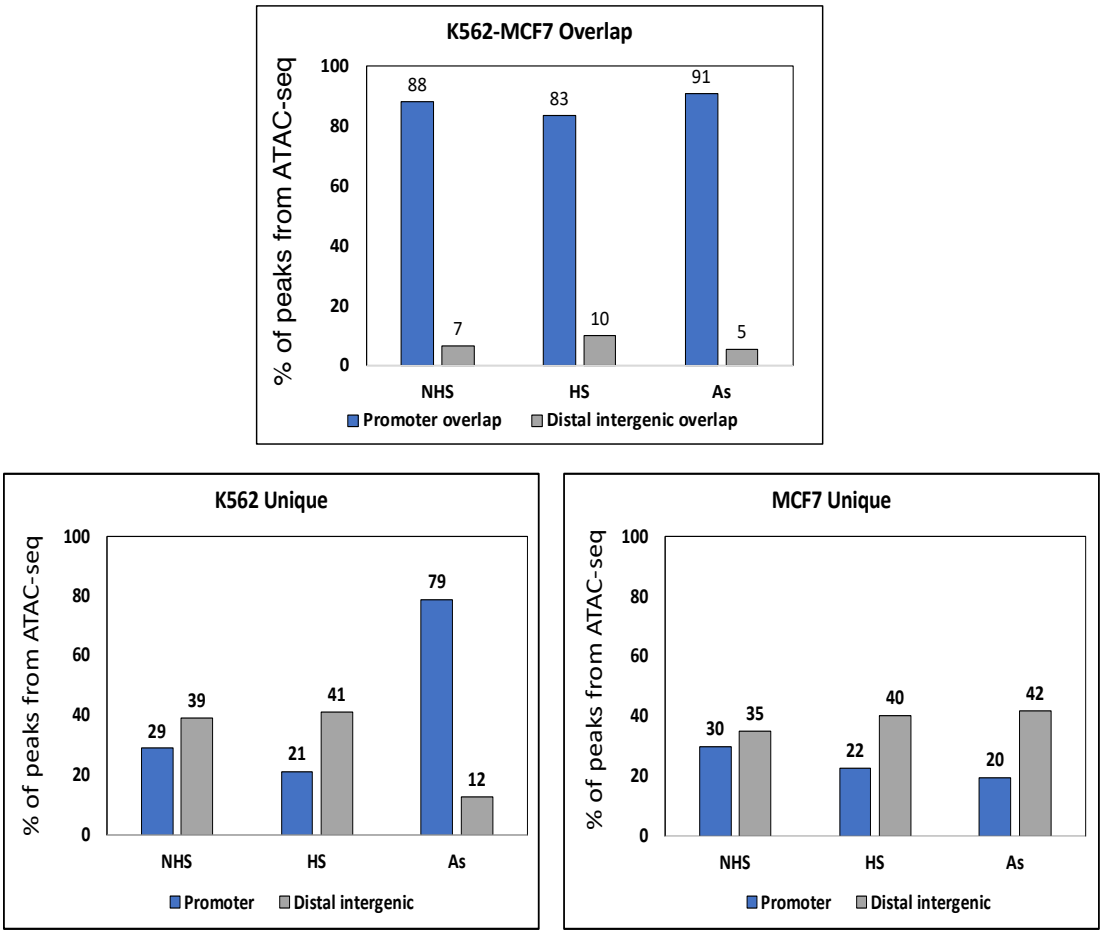


Figure 3.35. Similarity in ATAC-seq peaks between two cell types arise from promoters. Barplot showing overlapped regions between two cell types belong mostly to promoters (top) while unique peaks in both cell types arise from mostly distal and some promoters (bottom).

We further asked if ATAC peaks show different genome-wide distribution in the two cell types as HSF1 peaks. We merged all ATAC peaks that were called in both cell types and in all conditions from all replicates and created a master ATAC peaklist. We plotted heatmap for each of the individual ATAC samples against these peaklists and found there are distinct regions in each cell type that have unique ATAC peaks, similar to what we saw from our overlap analysis (Fig 3.36).

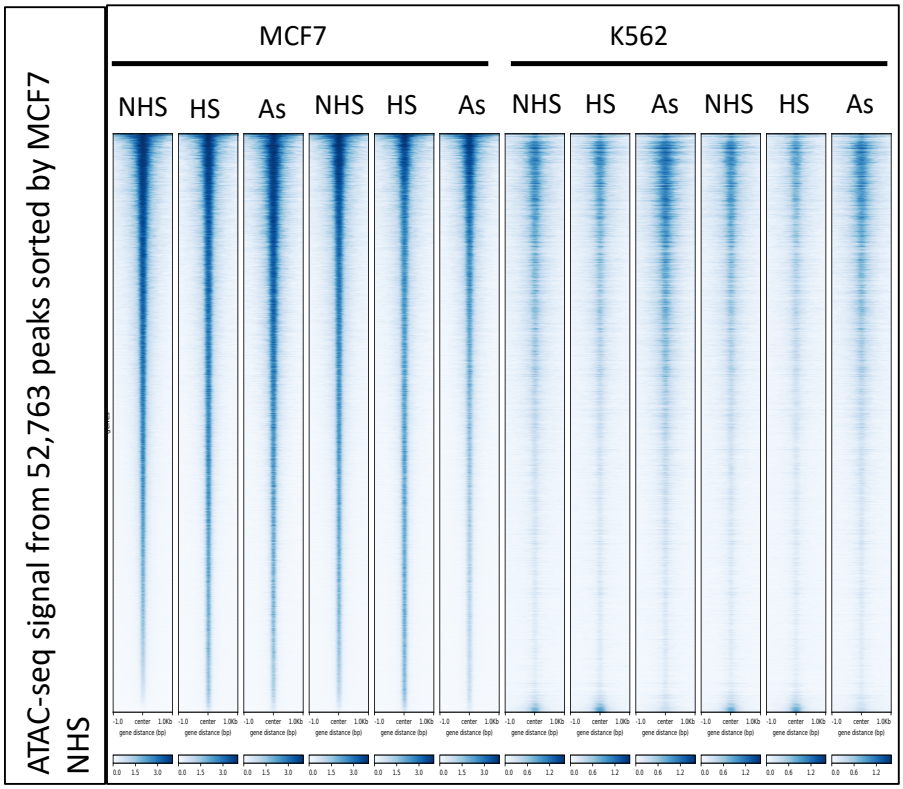


Figure 3.36. ATAC-seq signals are cell-type specific. Heatmap showing distinct distribution of ATAC-seq signal between MCF7 and K562 cell types; both replicates shown.

It is evident from the heatmap that ATAC signals in these two cell types show differences in distribution. Notably, the top ~1K peaks in the heatmap are more prominent in MCF7 while the bottom ~1K peaks have stronger signals in K562. We annotated these peaks that are distinctly different in these two cell types and found the top 1k peaks are mostly equally distributed between promoters and distal regions whereas the bottom 1k peaks are almost exclusively distal peaks (Fig 3.37).

We were curious if these distal intergenic regions contain any enhancers or super-enhancers. We overlapped these regions with cell-type specific typical enhancers (TE) and super-enhancers (SE) downloaded from the latest curated super-enhancer database (http://www.licpathway.net/sedb/). We found that top 1k peaks are more enriched in MCF7 cell-specific enhancers than K562, whereas the bottom 1k peaks have only one SE and 35 TE in MCF7 cells whereas K562 cells have 55 SE and 364 TE (Fig 3.38). This result show that there are enhancers and super enhancers that can be detected by ATAC signals and are specific to each cell type.

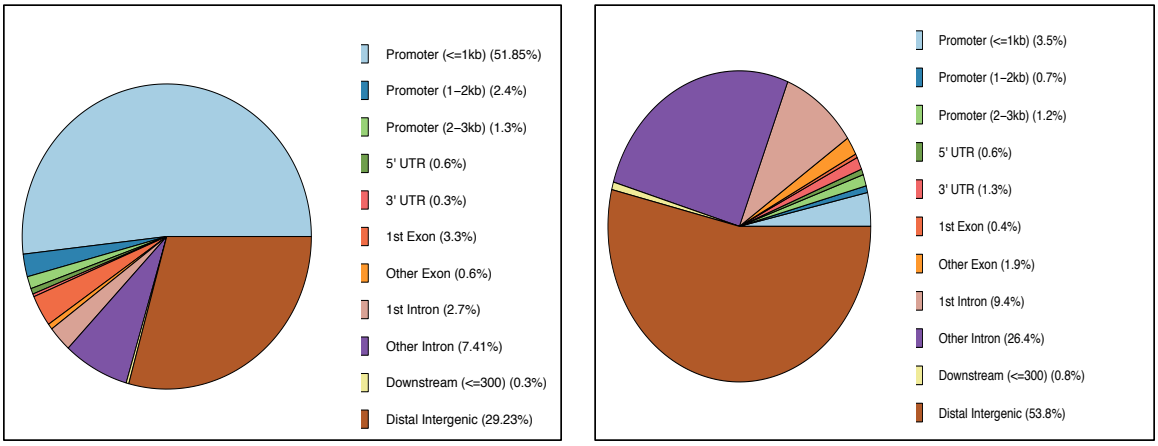


Figure 3.37. ATAC-seq signals from unique peaks arise from promoter-distal regions. Top 1000 and bottom 1000 ATAC-peaks annotation showing promoter enrichment (left) and enrichment of signal at distal-intergenic region (right).

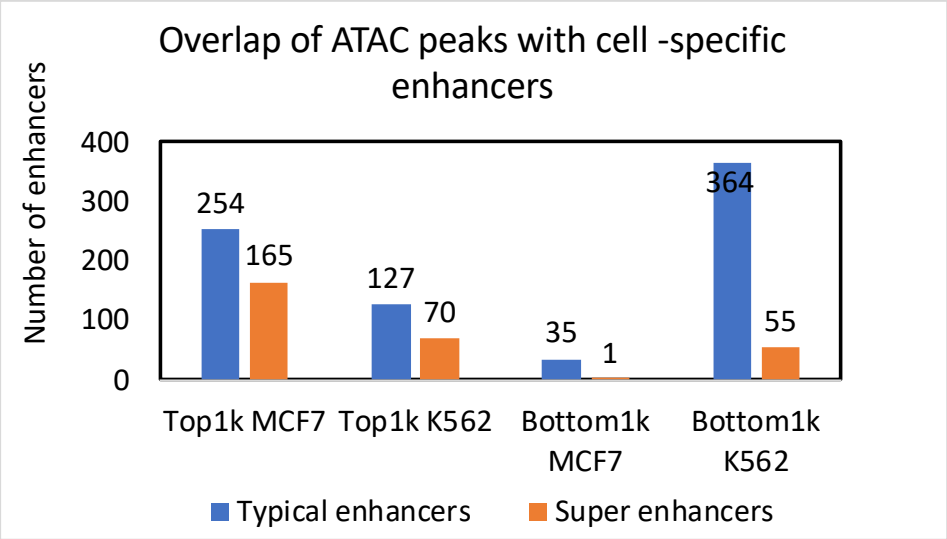


Figure 3.38. Unique ATAC peaks are enriched in cell-specific enhancers. Bar plot showing distribution of typical enhancers and super enhancers among the top 1000 peaks and bottom 1000 peaks in each cell type

From the UCSC browser shot (Fig 3.39; left) of one of the predicted SEs in K562 cells at the bottom 1k peaks, it is evident that these enhancers show two distinct states in two different cells – while this region is open and accessible in K562 cells, they are closed in MCF7 cells. Furthermore, these SEs also respond to stimuli differently – with HS the openness is not affected much however with As there is a decrease in peak height suggesting that these enhancers might become less accessible with heavy-metal stress. UCSC browser shot of MCF7 specific SEs further confirms the observation the presence of open cell specific enhancers in MCF7, however the same region is closed in K562 (Fig 3.39; right).

We further asked how many of these cell-specific enhancers are common between these two cell types and if any correspond to any of the common pool of enhancers that are differentially activated in these two cells. We found there are 4000 TEs and 64 SEs that are common between MCF7 and K562 cell types. When we asked if any of the top or

bottom 1k ATAC peaks correspond to these common enhancers we found there are only 2% TEs and 0.2% SEs in the top and 0.15% TEs and no SEs in bottom 1k peaks (Fig 3.40).

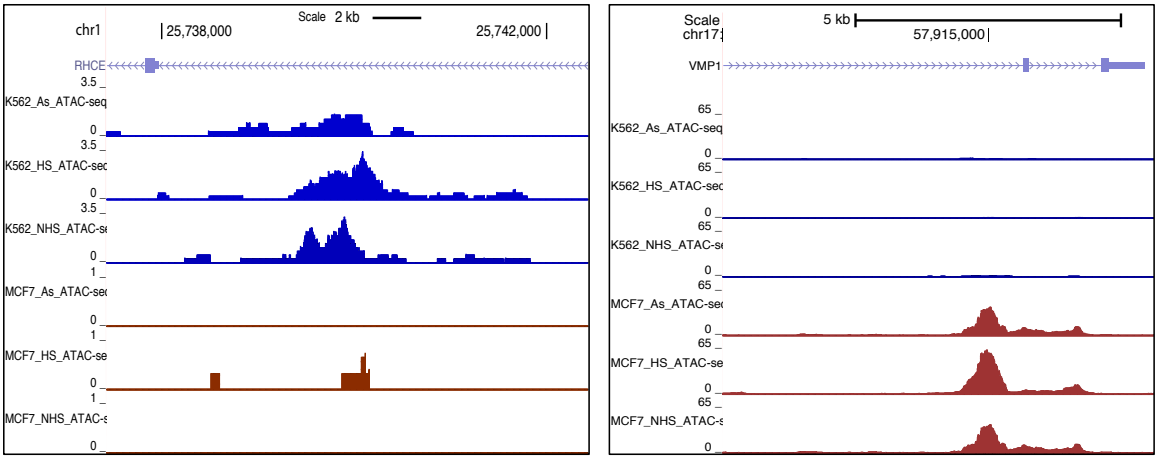


Figure 3.39. UCSC browser shot showing K562 cell-specific super-enhancer (left) and MCF7-specific super-enhancer (right) showing chromatin openness in only the specific cell-type whereas is a closed chromatin region in the other cell-type.

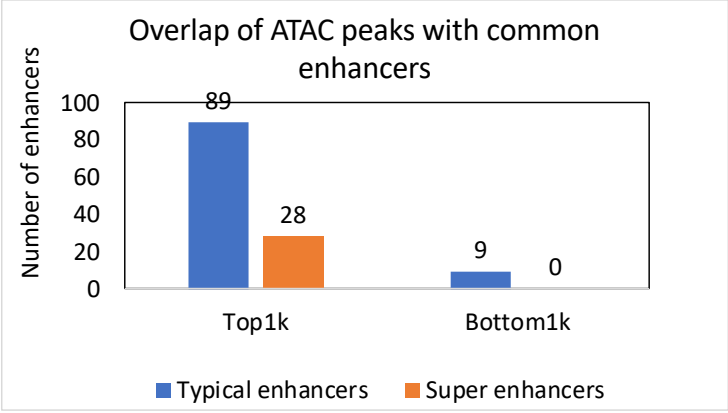


Figure 3.40. Super-enhancers are cell-type specific and do not show significant overlap between the two cell types. Barplot showing distribution of common enhancers among unique MCF7 and K562 peaks is negligible.

UCSC screenshot of a super-enhancer that is present in both MCF7 and K562 show that the enhancer is more accessible in MCF7 cells than in K562 indicating again that same enhancers can also be present in different states in two different cell types (Fig 3.41).

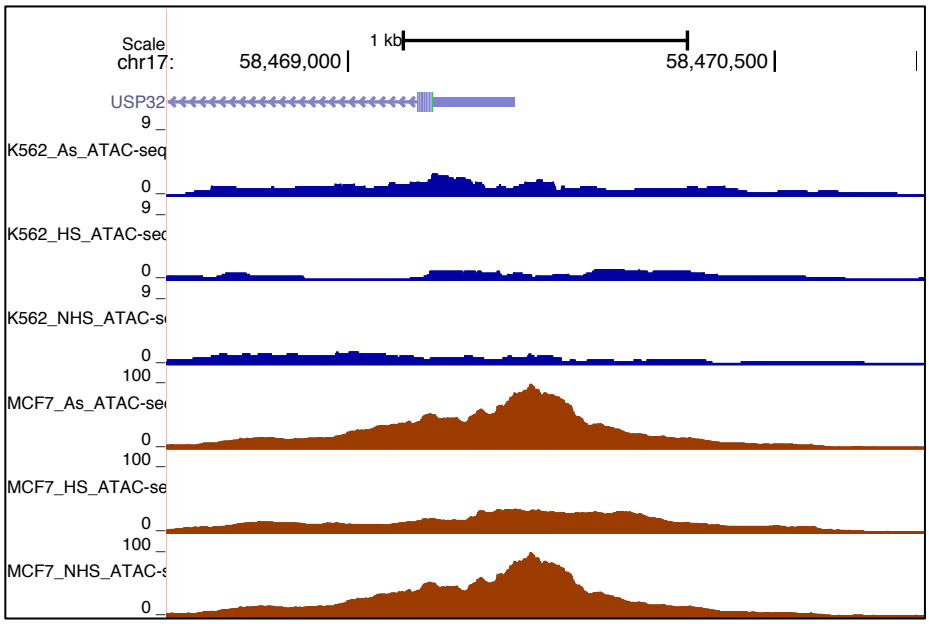


Figure 3.41. Predicted super-enhancers exhibit different states of activity in different cell types. UCSC browser shot of a super-enhancer that is present in both MCF7 and K562 but is only accessible in MCF7 cells.

Overall this study shows that different cells utilize different HSR programs when exposed to the same stress and these differences are more likely to originate from enhancers rather than promoters. Further studies are needed to address more in-depth questions about the dynamic states of enhancers in these two cell types and to identify the distinction between a predicted SE that is not accessible versus an unpredicted SE. Connecting enhancer states with histone marks and transcription might answer some of these questions on poised versus active enhancers.

CHAPTER 4

TRANSCRIPTIONAL DYNAMICS OF POL II DURING HEAT SHOCK RESPONSE (HSR) REVEAL DISTINCT MECHANISMS OF REGULATION IN DIFFERENT CELL TYPES

4.1 Functional States of Genes during HSR

Heat shock response (HSR) is characterized by three classic features – (i) genome-wide HSF1 binding at multiple loci, (ii) activation of HSR genes that maintain homeostasis in cells during stress and, (iii) genome wide transcriptional repression. In 1980, Susan Lindquist noted that if *Drosophila* cells were exposed to a range of high temperatures, those cells completely shut down protein synthesis except for a few HS-responsive genes such as hsp70, and undergo massive transcriptional repression (Lindquist, 1980) (Fig 4.1)

Recently with the advent of new techniques to measure nascent RNA transcription such as GRO-seq, NET-seq, and PRO-seq, multiple studies have been performed to uncover Pol II dynamics in mammalian cells and its relationship with gene expression. Unlike Drosophila, mammalian transcription at elevated temperature continues largely unabated across the genome[103] (Fig 3.7A). Transcriptional repression of genes in HS, while extensive, affects only a proportion of active genes (~ 49% of active genes and ~34% of all genes according to Mahat et al., 2017) that is typical of a normal response to stimuli such as hormones (50% downregulation of active genes in response to estrogen, Hah *et al*, 2011). These and other studies have revealed an unexpected complexity of the seemingly well-studied HSR.

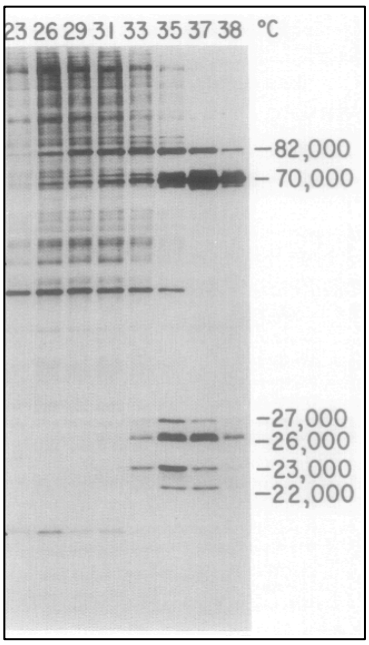


Figure 4. 1 Massive repression in Drosophila cells on exposure to heat. Gel electrophoresis of Schneider 2 cells of Drosophila exposed to different temperatures for 1h and labeled with [^{3}H]leucine showing massive repression of proteins with increased heat except for heat-shock proteins (molecular weight around 80KD). Picture reused from *Lindquist et al, Developmental Biology, Volume 77, Issue 2, 15 June 1980, Pages 463-479* with permission.

The mechanisms of transcriptional repression are still debatable. According to one school of thought regulation happens during Pol II recruitment at promoter; during stress, recruitment is blocked and hence there is repression. In contrast, another hypothesis suggests repression is a result of increased Pol II pausing and decreased release of Pol II into gene bodies (Fig 4.2).

There are also controversies regarding how Pol II exists in the cell. According to Hieda *et al*, 2005, in Chinese hamster ovary cells, there are three distinct populations of Pol II – (i) a hypo-phosphorylated form that is mobile and freely available for recruitment, (ii) a hyper-phosphorylated form that is bound to promoter and dissociates on repression by heat stress, and (iii) an engaged form of Pol II that dissociates when initiation and

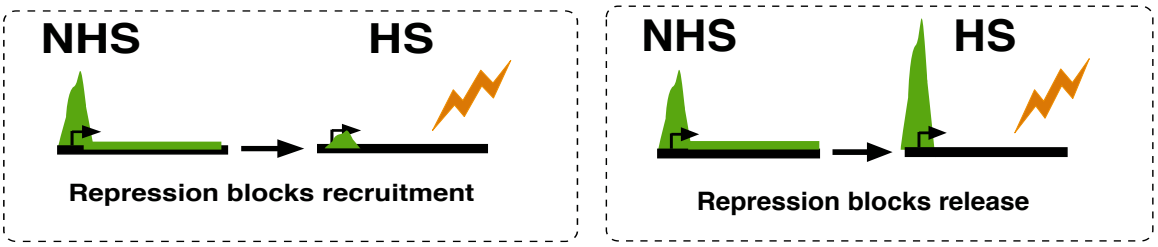


Figure 4. 2. Two possible models of gene repression: by blocking Pol II recruitment at promoter (left) and by blocking Pol II release onto gene body (right) with stress.

elongation phase of transcription is blocked. They show with imaging techniques such as FRET that upon heat shock at 42°C, Pol II is dissociated from chromatin and this dissociation of Pol II away from promoters leads to repression of transcription. This dissociation is transient and steady-state transcription is recovered after incubating the heat-shocked cells for 4.5h at physiological temperature[112].

In this section of my dissertation, I would address the mechanisms of genome-wide transcriptional changes in response to HS in MCF7 cells. To understand if same cells use different mechanisms of transcriptional response in different stresses inducing HSR pathway, I would expose MCF7 cells to Arsenic (As) and measure the effects using PRO-seq and ChIP-seq against Pol II. To further dissect transcriptional mechanisms, we asked if a completely different cell type is exposed to the same stress inducing HSR, will the mechanisms of gene activation or repression be the same or change with cell type.

4.2 MCF7 Cells Do Not Show Evidence of Massive Repression upon HS or As

To determine if MCF7 cells exhibit a massive repression such as is observed in *Drosophila*, we performed nascent run-on transcription experiments on MCF7 at different time points of HS treatment: 12 min, 30 min, and 60 min to detect global changes in nascent

transcription. PRO-seq analyses of MCF7 cells reveal that there is no massive repression of genes upon HS treatment, as was previously observed by Mahat et al., 2016 in mouse fibroblast cells. Only 27 genes (0.0001%) and 3779 genes (0.16%) were found to be repressed at 12 min, and at 30 min, respectively. However, unlike what has been reported by Mahat et al., 2016 that increased periods of HS led to increased gene repression, we saw with 60 min of HS treatment, the number of repressed genes goes down to 767 from 3779.

This result contradicts the massive genome-wide repression of non heat-shock genes observed in *Drosophila*. To determine if this lack of massive repression is a HS-specific event, we used a HS-independent stimulus and exposed MCF7 cells to 1h of 500μM As at ambient temperature. Exposure to As led to 5.3 times more repressed genes at 60 min than HS in MCF7 cells. However, majority of the genes remain unaffected (Fig 4.3).

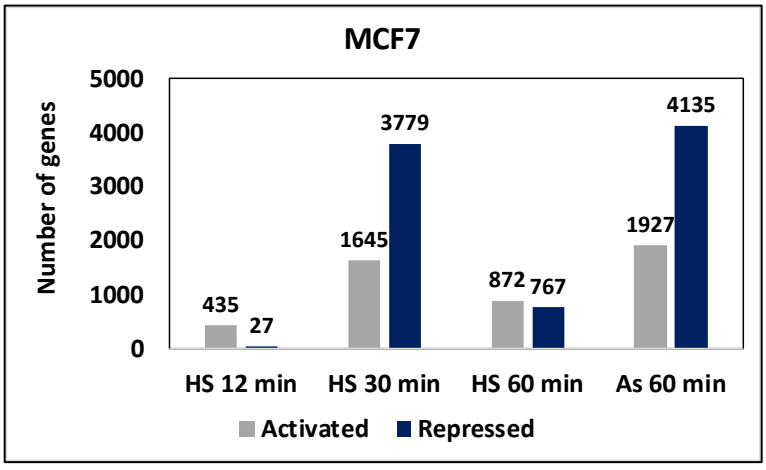


Figure 4. 3. MCF7 cells do not undergo massive repression in response to HS or As. PRO-seq analysis of gene body counts using DESeq2 pipeline ($p < 0.01$) showing number of activated and repressed genes in MCF7 cells treated with HS and As.

4.3 Validation of HS System in Published Mammalian HSR Model: K562 Cells

To determine differences in number of repressed genes and the extent of repression is not due to experiments having been done in different laboratories and could possibly be technical, we used K562 cell line which is a validated model system for HS and exposed them to 60 min of HS or 60 min of As. RT- qPCR analysis of HS-target genes in K562 show upregulation of mRNA with HS or As. Although with As in K562 cells, response is less robust (Fig 4.4).

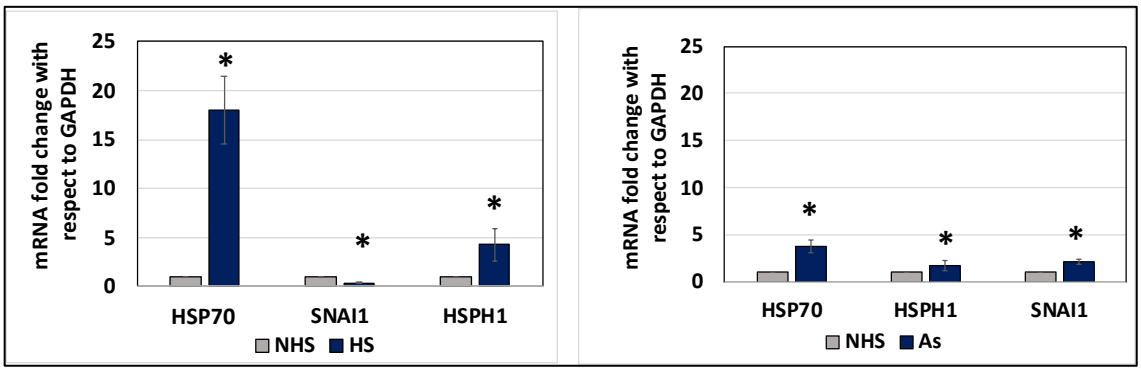


Figure 4. 4. K562 cells respond to HS or As. RT-qPCR results showing K562 cells respond to both HS (left) and As (right), similar to MCF7 cells.

PRO-seq analysis on our K562 cells reveal that these cells undergo greater repression of gene expression on exposure to HS or As than those of MCF7 cells. The numbers of activated and repressed genes upon HS in K562 cells are comparable with previously published data in these cells (Fig 4.5) and validate our observed differences in the extent of gene repression between the two different cell lines.

This validates the HS treatment performed in our lab and that the lack of repressed genes in MCF7 cells is not an effect of inadequate HS treatment.

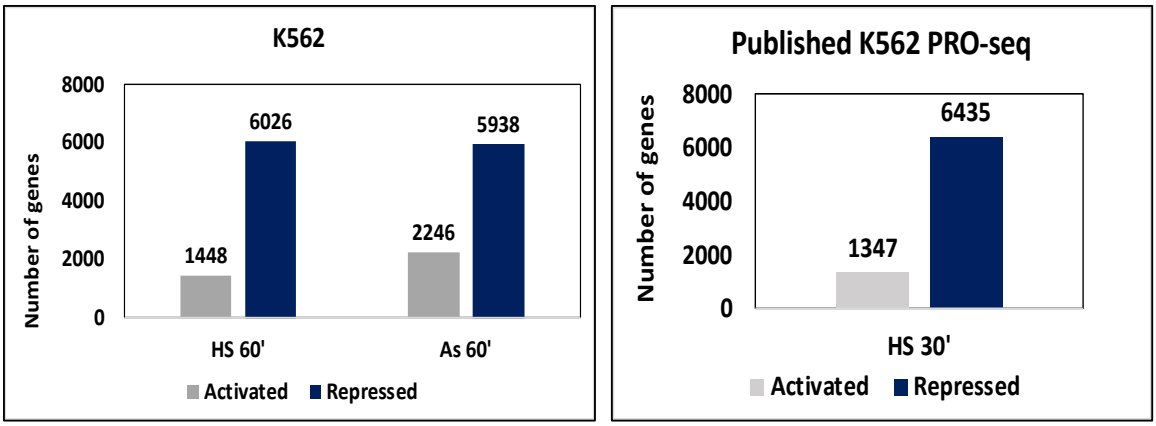


Figure 4. 5. K562 HS data is comparable to published data confirming validity of heat-shock treatment. PRO-seq analysis of gene body counts using DESeq2 pipeline ($p < 0.01$) showing number of activated and repressed genes in K562 treated with HS and As in lab (left); (right) published K562 PRO-seq data on HS treated cells analyzed using the same pipeline as ours showing similar numbers of activated and repressed genes.

4.4 Difference in Number of Repressed Genes in K562 Versus MCF7 Cells Is Not Due to Difference in the Number of Active Genes at Ground State

Since K562 demonstrated 7.3 times more repressed genes than MCF7, we were curious if this difference in repression between the two cell types is an effect of different number of active genes at ground states of the cells. To test this possibility, we categorized genes that had less than 1 read per kb density at gene body regions as inactive genes and genes that had more than 1 read per kb density at gene body regions as active genes. As expected, we found the number of active genes and inactive genes are very similar in both cell types at ground state (Fig 4.6).

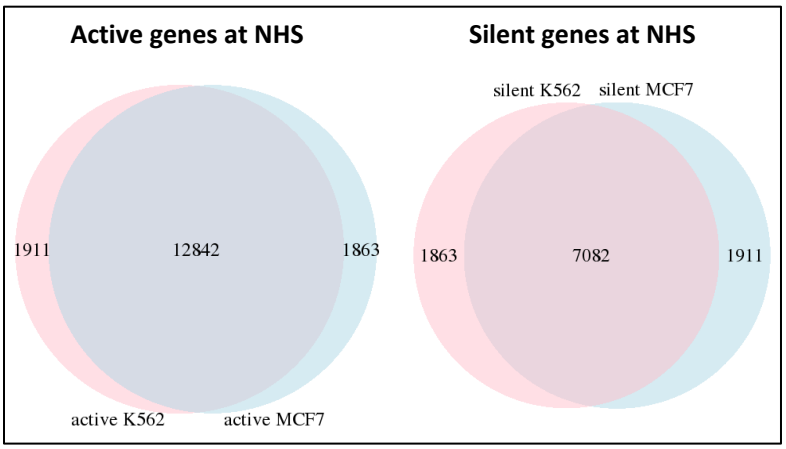


Figure 4. 6. Active and inactive genes at ground state between the two cell lines are similar in number. Venn diagram showing gene overlap between active genes (left) and inactive genes (right) in MCF7 and K562 cells.

4.5 Majority of the Activated and Repressed Genes Overlap between the Two Cell Types

To investigate if HS or As affect different genes in the two cell types, an overlap was done using VennDetail Shiny App (http://hurlab.med.und.edu:3838/VennDetail/). 54% of genes that are upregulated and 53% of genes that are downregulated in HS in MCF7 are common to K562. The numbers are similar also in case of As response (Fig 4.7). This is consistent with what we observed with respect to HSF1 binding at promoters between the two cell types. These results indicate that although the two cells differ in the number of genes affected, there is a core group of HS response genes that, irrespective of cell type, will respond similarly consistent with their dependence of HSF1 binding. 70% of HS-overlapped genes and 63% of As-overlapped genes show HSF1 binding upon upregulation of gene expression with the treatments.

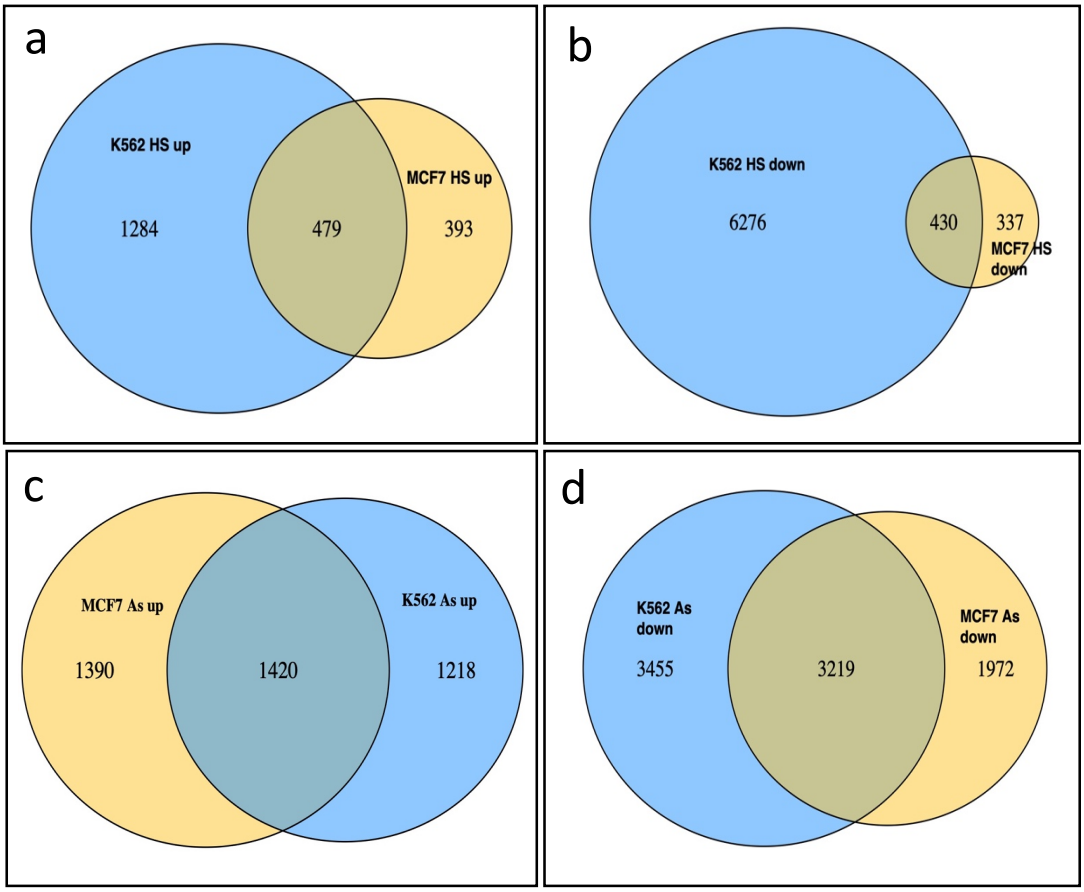


Figure 4. 7. Both cell types exhibit a common pool of activated and repressed genes. Venn diagram showing overlap between different cell types MCF7 and K562: HS-upregulated genes (a), HS-downregulated genes (b), As-upregulated genes (c), and As-downregulated genes (d).

Interestingly when we overlap genes that are upregulated or downregulated in HS and As, there is an overlap of more than 50% genes in MCF7 cells in both cases. However, although the number of downregulated genes common between HS and As in K562 is more than 66%, the number of upregulated genes is less than 35% (Fig 4.8). These results indicate that the genes that are repressed in HSR are very similar between different stressed conditions, however the genes that are activated might be stimulus specific.

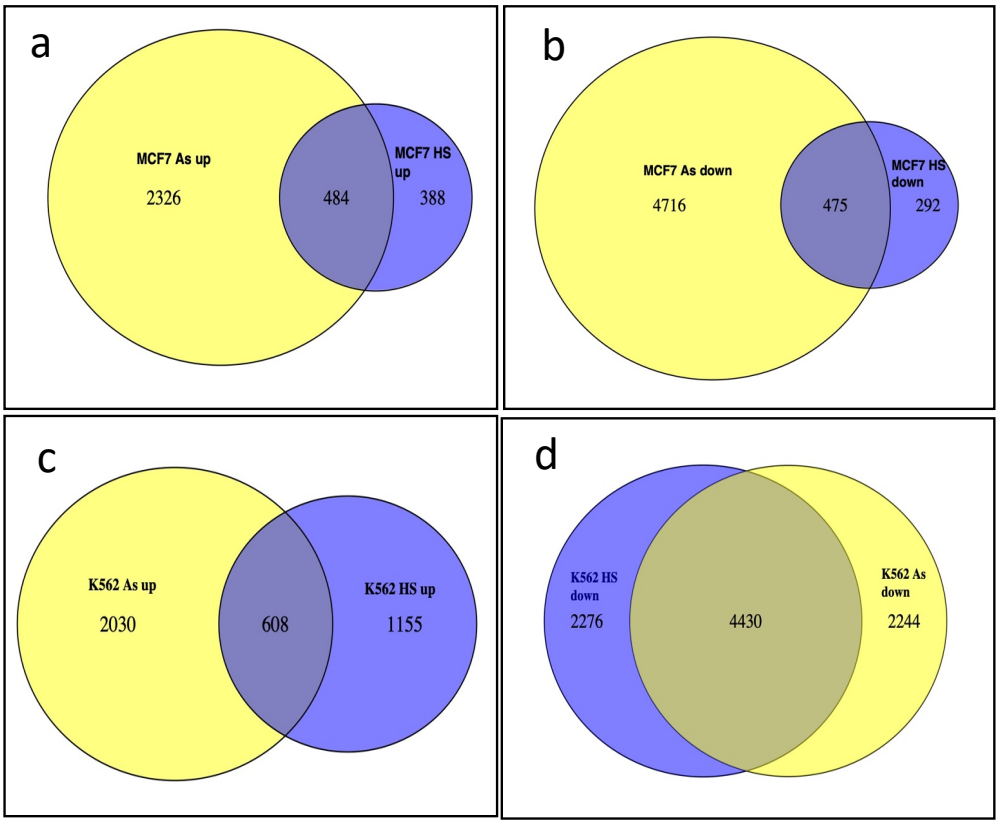


Figure 4. 8. K562 shows greater difference in common genes upregulated in HS and As than MCF7. Venn diagram showing overlap between different treatments HS and As subjected to either MCF7 or K562: MCF7 upregulated genes (a), MCF7 downregulated genes (b), K562 upregulated genes (c), and K562 downregulated genes (d)

4.6 Stress Response is Mediated through a Common Cohort of Genes Conserved Across Distinct Cell Lines and in Distinct Stresses

We next asked which functional pathways do the genes that are common, and unique in different cell types and different conditions belong (Fig 4.4 and Fig 4.5). To address that, we used Bioconductor packages and Reactome Pathway database, and characterized the gene-lists using 3 different GO categories – Cellular Components (CC), Biological Functions (BP), Molecular Functions (MF), and KEGG pathway analysis. Only the pathways that represented the maximum hits for the gene-lists are shown in the figures.

A majority of genes that show upregulation in MCF7 and K562 cell lines with HS treatment are the ones involved in temperature and heat sensing, heat acclimation, genes involved in protein-folding and genes involved in estrogen signaling pathway (Fig 4.9).

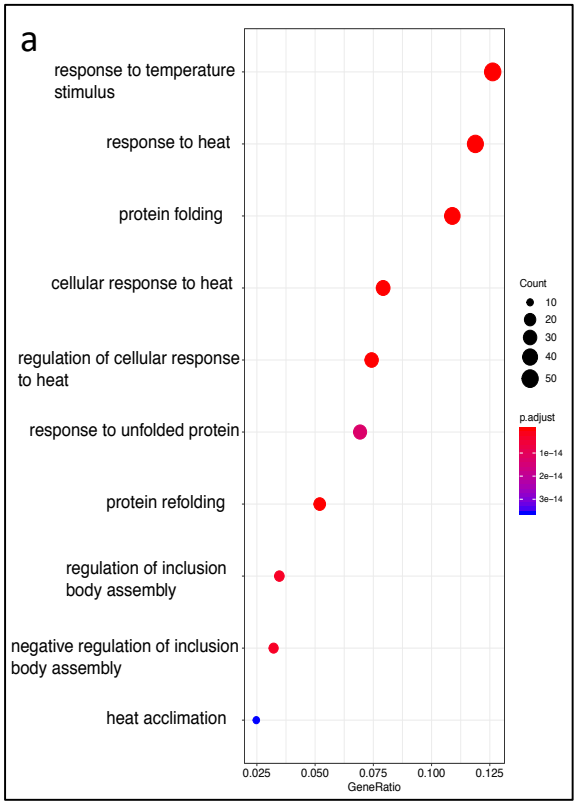
Similarly, with As treatment, genes that are involved with heat-shock sensing, protein folding, and MAPK pathway are upregulated (Fig 4.10). Rapid upregulation of MAPK pathway with acute As stress has also been observed previously in immortalized human keratinocyte cells (Cooper *et al*, 2004).

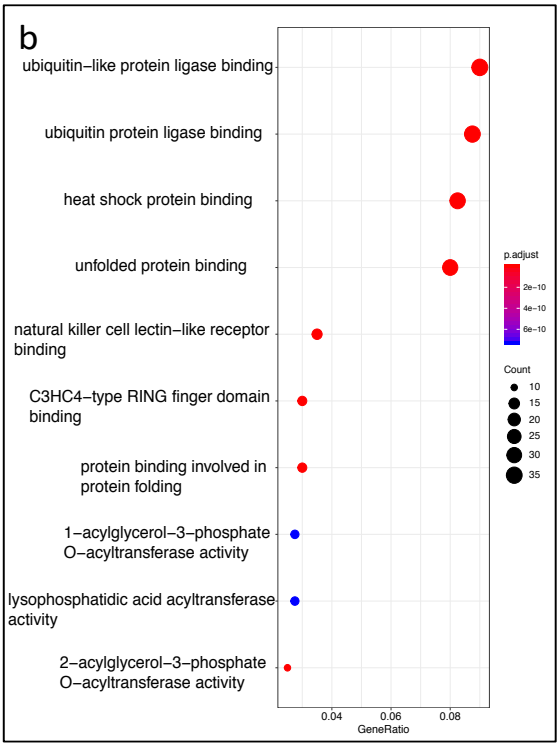
Downregulated genes in both cell types in response to HS include genes that are responsible for cell-cycle transitions, DNA replication and chromatin binding indicating that under stress these pathways are universally shut down to maintain fidelity of replication in cells (Fig 4.11).

Downregulated genes in both cell types in response to As include gene that are responsible for development and maintenance of cellular systems and downregulation of genes involved in p53 pathway. Genes in p53 are known to as tumor-suppressors and exposure to As can lead to tumorigenic effects by blocking that pathway (Fig 4.12). As is already known to be a potent carcinogen by International Agency of Cancer Research (IACR).

Uniquely activated genes in MCF7 with HS are involved in blocking activation of pathways by negative regulation of phosphorylation and in RNA localization whereas in K562, unique genes are mostly involved in metabolic processes (Fig 4.13).

Figure 4. 9. HS activates a core gene network involving genes responsible for sensing temperature, regulating protein misfolding and genes in estrogen signaling pathway. Pathway analysis of common HS-upregulated genes between MCF7 and K562;
(a) Biological Processes involved
(b) Molecular Functions involved
(c) KEGG pathway analysis


a
response to temperature stimulus
response to heat
protein folding
cellular response to heat
regulation of cellular response to heat
response to unfolded protein
protein refolding
regulation of inclusion body assembly
negative regulation of inclusion body assembly
heat acclimation
0.025
0.050
0.075
0.100
0.125
GeneRatio
Count
10
20
30
40
50
p.adjust
1e−14
2e−14
3e−14


b
ubiquitin−like protein ligase binding
ubiquitin protein ligase binding
heat shock protein binding
unfolded protein binding
natural killer cell lectin−like receptor binding
C3HC4−type RING finger domain binding
protein binding involved in protein folding
1−acylglycerol−3−phosphate O−acyltransferase activity
lysophosphatidic acid acyltransferase activity
2−acylglycerol−3−phosphate O−acyltransferase activity
0.04
0.06
0.08
GeneRatio
p.adjust
2e−10
4e−10
6e−10
Count
10
15
20
25
30
35


c
Protein processing in endoplasmic reticulum
Estrogen signaling pathway
Antigen processing and presentation
Legionellosis
Longevity regulating pathway − multiple species
Arginine and proline metabolism
Prion diseases
0.04
0.06
0.08
0.10
GeneRatio
Count
5
10
15
p.adjust
0.01
0.02
0.03

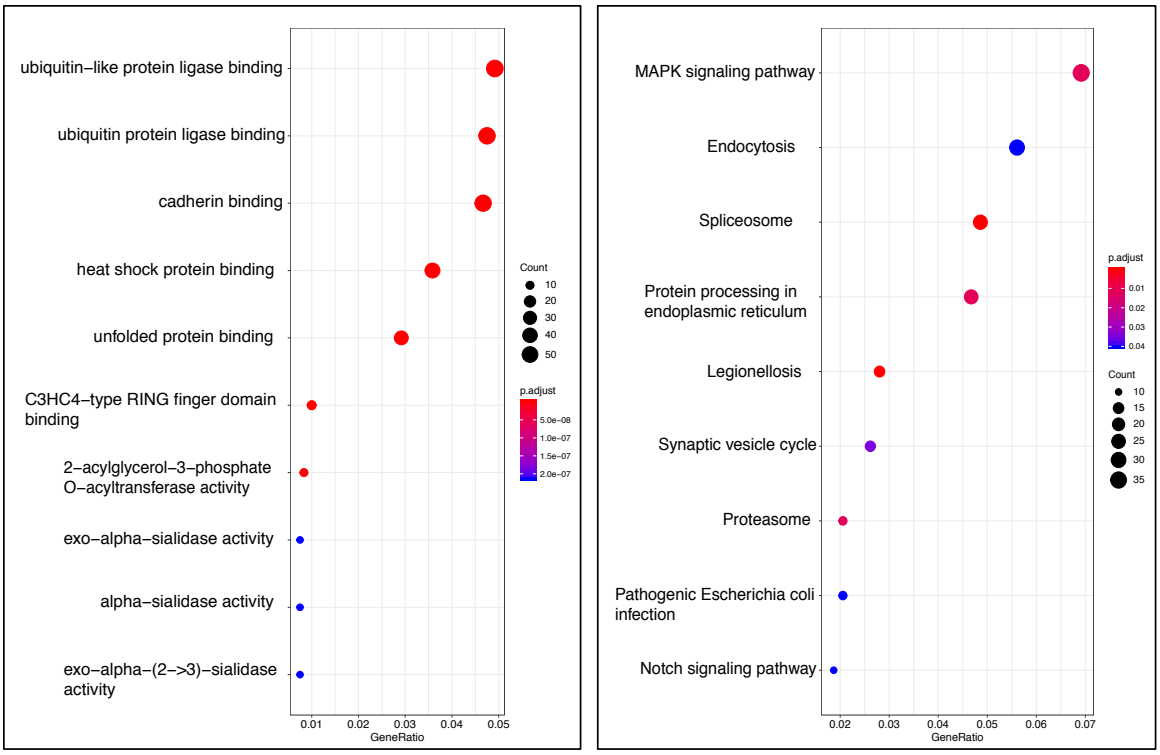


Figure 4. 10. As activates heat-shock proteins, stress-sensing pathways and MAPK pathway in both cells. Pathway analysis of common As-upregulated genes between MCF7 and K562; (left) Molecular Functions involved (right) KEGG pathway analysis.

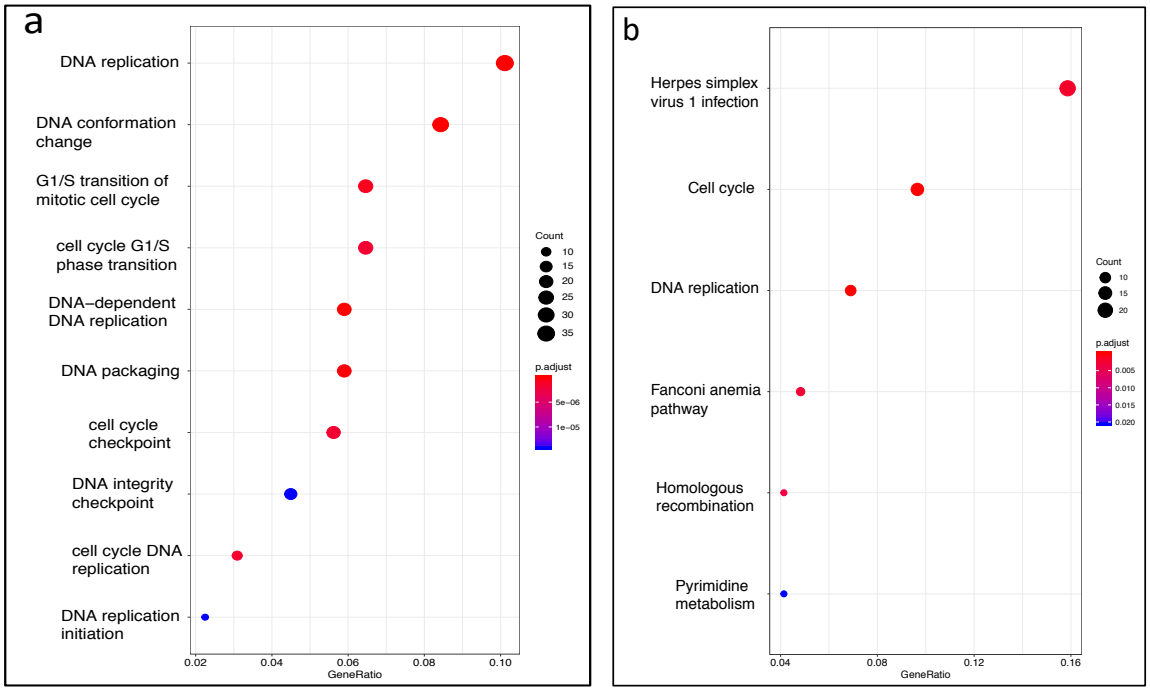


Figure 4. 11. Both cells shut down processes that involving DNA replication and cell-cycle transition possibly to maintain fidelity in cells. Pathway analysis of common HS-downregulated genes between MCF7 and K562;
(a) Biological processes involved
(b) KEGG pathway.

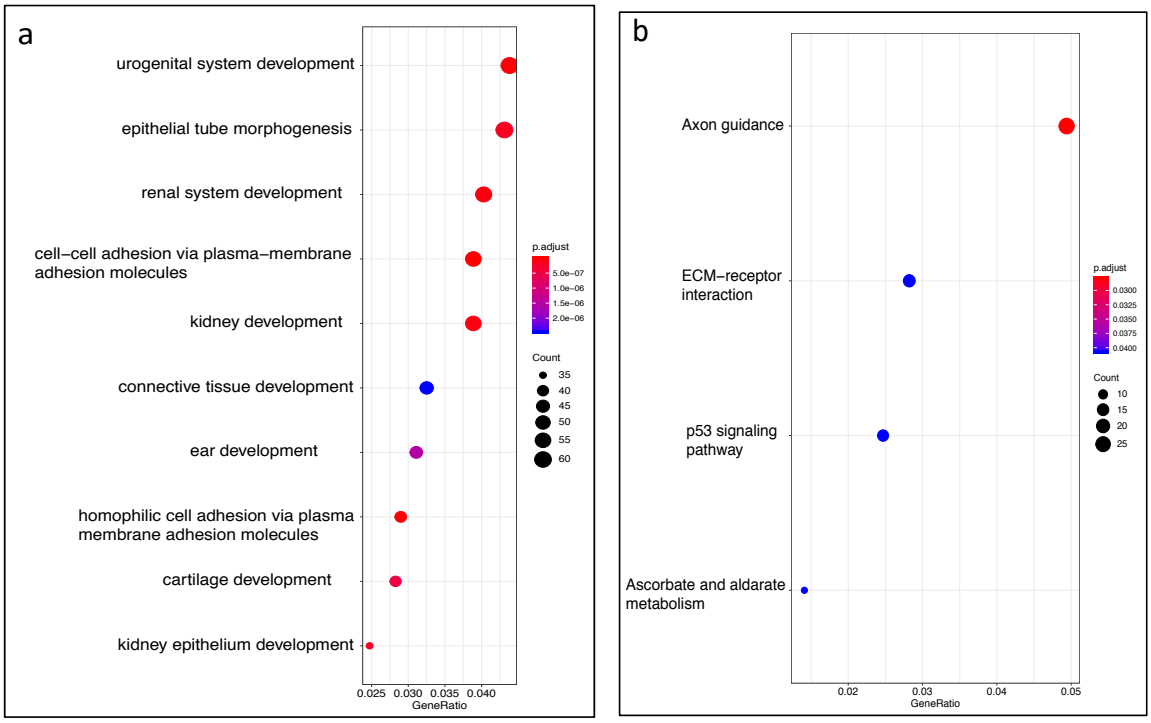


Figure 4. 12. As exposure leads to downregulation of p53 pathway in both cells, hence exerting its carcinogenic activity. Pathway analysis of common As-downregulated genes between MCF7 and K562;
(a) Biological processes involved
(b) KEGG pathway analysis

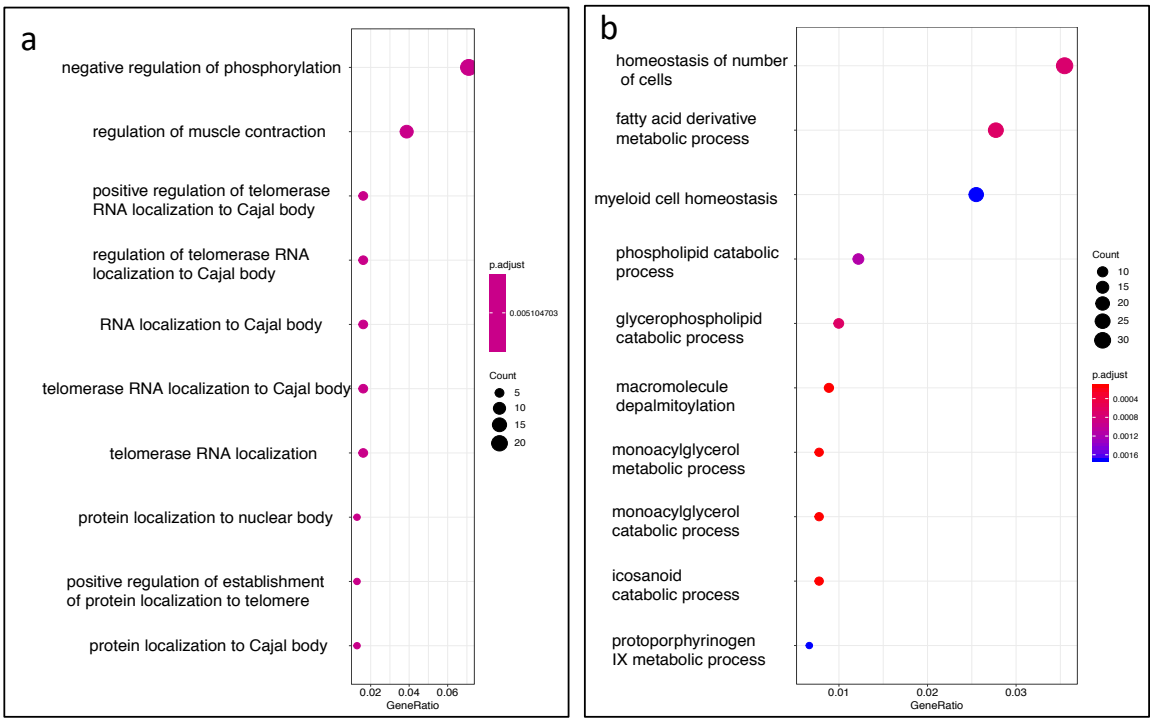


Figure 4. 13. Uniquely activated genes in MCF7 (a) and K562 (b) with HS are responsible for maintaining RNA fidelity and regulating different metabolic processes respectively. Pathway analysis of unique HS-upregulated genes in MCF7(a) and K562 (b).

Uniquely activated genes in MCF7 with As are involved cytoskeletal organizations and inflammation (Fig 4.14 a) whereas in K562, unique genes are involved in inflammatory processes (Fig 4.14 b).

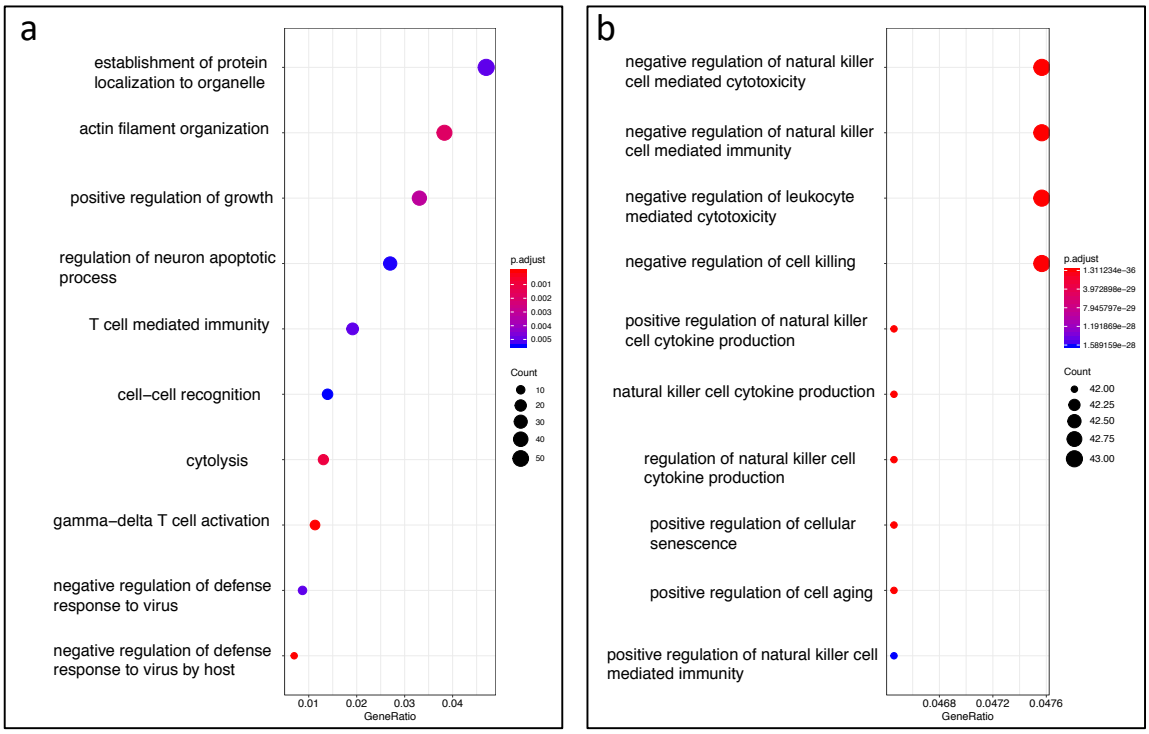


Figure 4. 14. Uniquely activated genes in MCF7 (a) and K562 (b) with As are responsible for triggering various inflammatory pathways. Pathway analysis of unique As-upregulated genes in MCF7(a) and K562 (b).

Uniquely repressed genes in MCF7 (Fig 4.15 a) and in K562 (Fig 4.15 b) with HS are involved in regulating various metabolic processes and RNA processing.

These results show that there is a common cohort of genes that are conserved across cell lines and treatments and respond to similar stimuli in the same way to protect cells from stress. The genes that vary between cell lines and treatments are the genes primarily involved in metabolic processes and cytoskeletal activity, which confirms Susan Lindquist's predictions (1986) where she postulates that the metabolic pathways are what determine cell specificity or stimulus specificity whereas the core-response genes remain same all across.

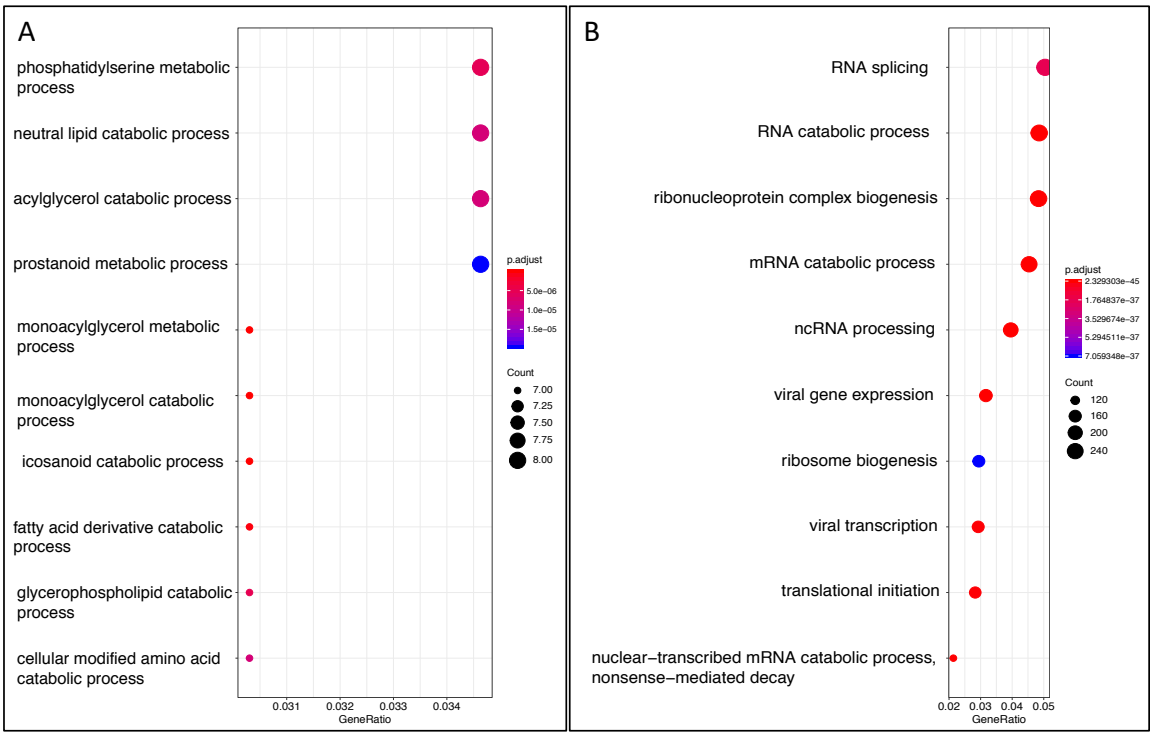


Figure 4. 15. Uniquely repressed genes in MCF7 (a) and K562 (b) with HS are responsible for cell-specific metabolic processes. Pathway analysis of unique HS-downregulated genes in MCF7(a) and K562 (b).

4.7 Mechanisms of Pol II Dynamics in HS or As Activated and Repressed Genes in MCF7 and K562

Previous results indicate that both cell types activate a group of HS-core response genes in stressed conditions and repress metabolic genes specific to cell type. As, which is a potent carcinogen, activates tumorigenic pathways and represses anti-tumor pathways. However, it is important to address whether these genes were activated or repressed in a similar fashion in both cells and in both conditions to understand the dynamics of gene regulation by Pol II.

4.8 Transcriptional Activation Occurs by Release of Pol II into Gene Bodies in Both Cell Types and Treatments

Most eukaryotic genes have bound Pol II at their promoter region[23]. These bound Pol II remain paused at promoter region and can either continue to make a full transcript or can abort transcription, depending on the stimulus. To address whether MCF7 or K562 cells utilize the same mechanism of activation in response to stress, we took genes that are activated either in HS or As in MCF7 and K562 and plotted their average signal at the promoter region from PRO-seq data. In both HS-treated MCF7 cells and K562 cells, paused Pol II is released from the promoter into the gene body resulting in activation of gene transcription (Fig 4.16). This mechanism of Pol II pause release has also been proposed by Lis group as a mode of gene activation and our results are consistent with available data.

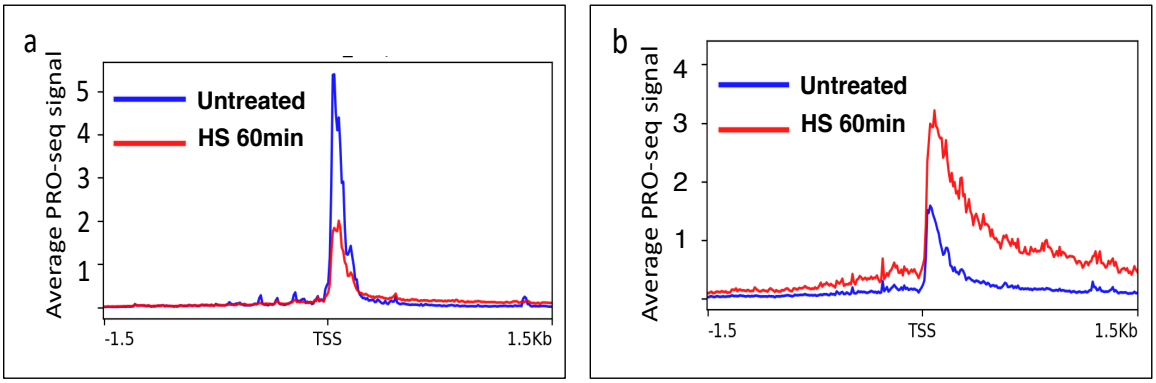


Figure 4. 16. Activation of gene transcription takes place through pause release. Profile of all HS-activated genes at NHS and HS condition showing release of Pol II in MCF7 (a) and K562 (b).

In As-treated MCF7 cells (Fig 4.17 a) or in K562 (Fig 4.17 b) cells under both stresses, there is an increase in Pol II recruitment at promoter and increased release of Pol II level to gene bodies.

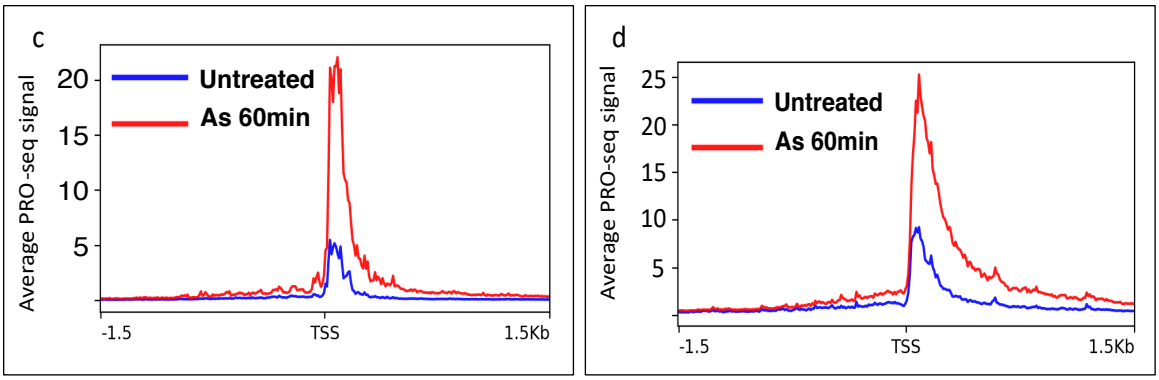


Figure 4. 17. Activation of gene transcription takes place through pause release in As. Profile of all As-activated genes at NHS and As condition showing release of Pol II in MCF7 (c) and K562 (d).

We visualized PRO-seq data (BigWig files) on UCSC genome browser and confirmed that both MCF7 and K562 activated genes show similar mechanisms of transcription activation. We looked at several HS-activated genes including DNAJB1 which has been previously shown[113] to be upregulated with HS in K562 cells (Fig 4.18).

These results indicate that the mechanisms by which gene activation take place are consistent between cell types and across treatments (Fig 4.19).

4.9 Transcriptional Repression Takes Place through Distinct Mechanisms in Different Cell Types

We next asked if repression of transcription in response to HS or As follows the same mechanisms in different types of cells. We used the same method for plotting average PRO-seq signal at promoters of repressed genes as we had used for activated genes. When we compared HS-treated MCF7 and K562 cells, we saw differences in Pol II regulation. In MCF7 cells, repression occurred by blocking recruitment of Pol II at promoter. In K562, recruitment appeared to be largely unaffected in repressed genes; however, there was less release of Pol II from promoter to gene bodies leading to accumulation of Pol II at promoters of repressed genes (Fig 4.20).

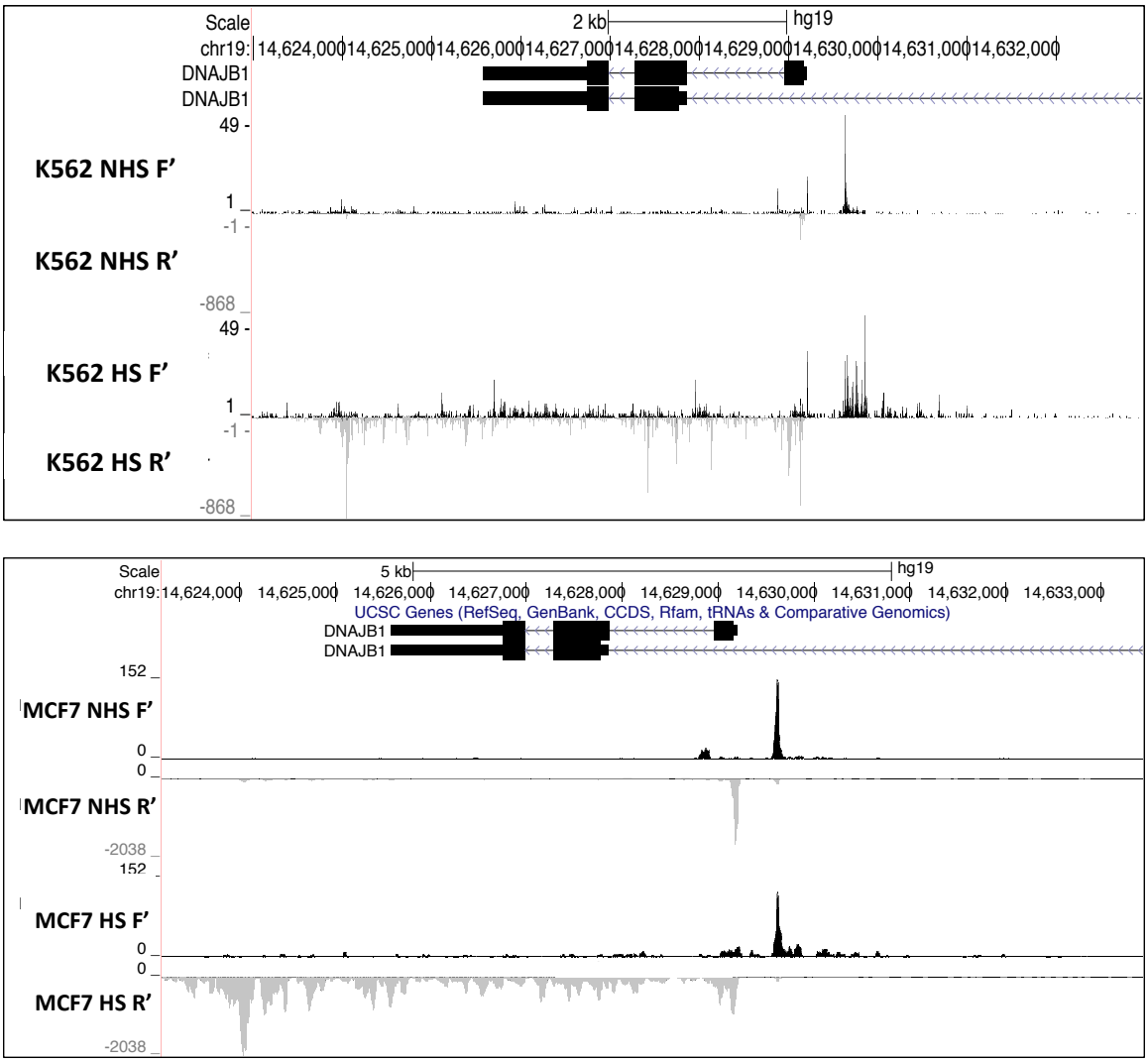


Figure 4. 18. Mechanism of activation of gene is conserved in both cell lines and in both treatments. UCSC browser shot of a HSR gene DNAJB1 that is shown in K562 (top) and MCF7 (bottom) to be activated by Pol II release in gene body.

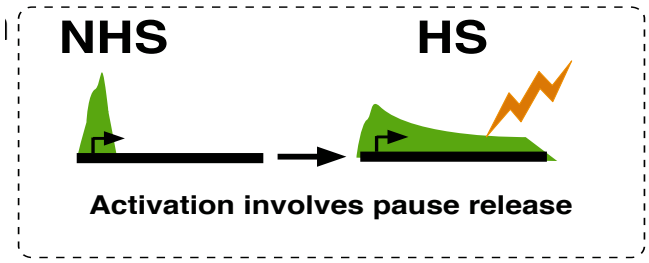


Figure 4. 19. Schematic showing the mechanism of gene activation in all conditions tested.

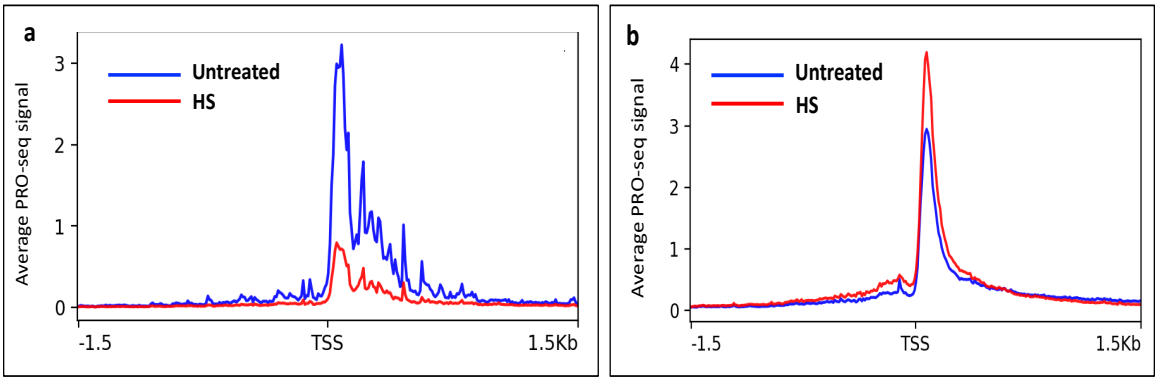


Figure 4. 20. Repression of gene transcription occur by different mechanisms in different cells. Profile of all HS-repressed genes at NHS and HS condition showing repression occurs by retention of paused Pol II in K562 (a) whereas in MCF7, recruitment of Pol II is decreased (b).

Similarly, when we compare As – treated MCF7 and K562 cells, we see similar differences in mechanisms between the two cells. However, the mechanisms of repression remain the same within the same cell type exposed to two different stimuli (Fig 4.21).

From these results we conclude that transcriptional activation may occur universally by releasing paused Pol II into gene body; however, mechanisms of transcriptional repression is distinct to cell type.

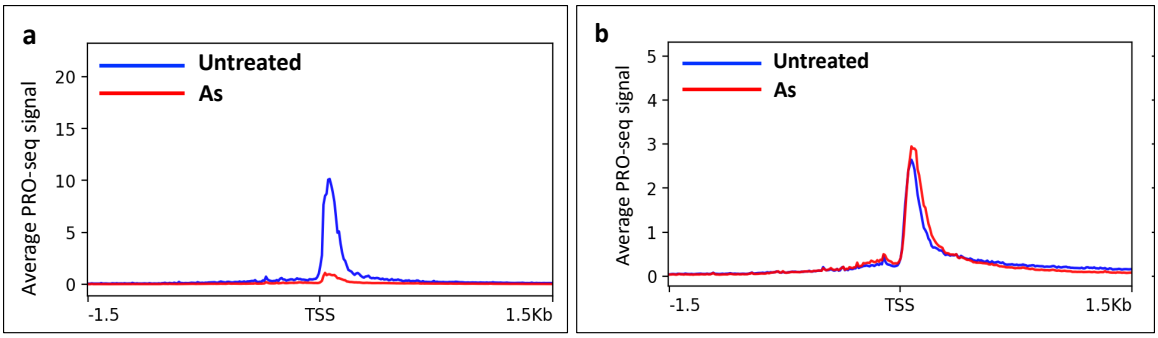


Figure 4. 21. Repression of gene transcription occur by different mechanisms in different cells with As stress. Profile of all As-repressed genes at NHS and As-treated condition showing difference in repression between MCF7 (a) and K562 (b).

4.10 HSR Genes Activated Both in MCF7 and K562 Show Variant Mechanisms of Transcription Repression

To determine if differences in mechanisms are driven by properties of genes or environment specific to the cell type, we took genes common to MCF7 and K562 that are either upregulated or downregulated for each of the treatments and generated PRO-seq read counts at promoter region. We plotted these overlapping differentially regulated genes (all DE genes), overlapping up-regulated and down-regulated genes for NHS, HS and As in MCF7 and K562 cells (Fig 4.22).

Both up-regulated and down-regulated genes have a higher or similar median value in HS than in NHS in K562 cells, indicating that there is accumulation of Pol II at promoters. Similarly, with As treatment in K562, upregulated genes show increase in Pol II level at promoter as expected while downregulated genes barely show any difference from their NHS states indicating decrease in release of paused Pol II from promoters (Fig 4.23).

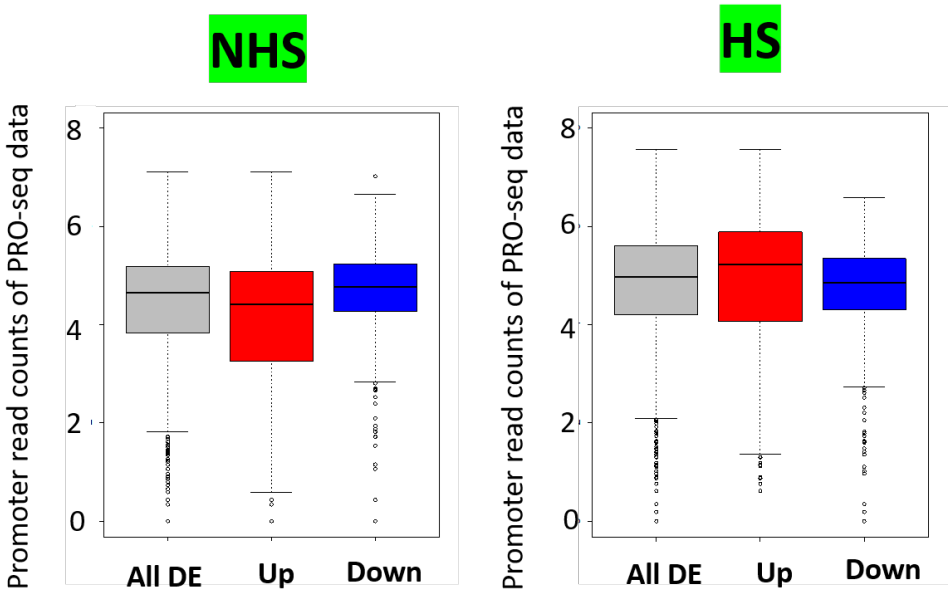


Figure 4. 22. Common downregulated genes between K562 and MCF7 show repression by retention of Pol II in K562 cells. Boxplots showing Pol II reads at promoter of common DE genes, up genes and down genes at NHS state (left) and with HS (right).

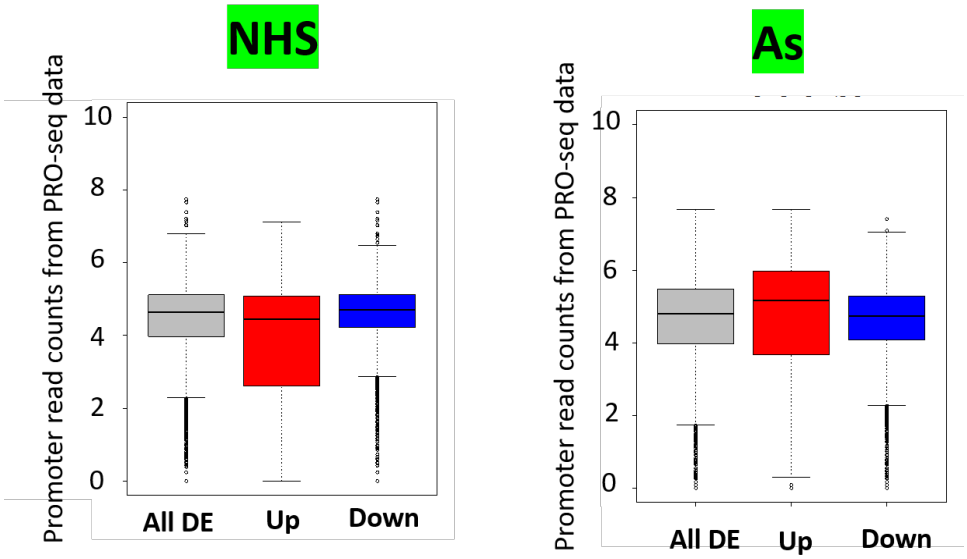


Figure 4. 23. Common downregulated genes between K562 and MCF7 show repression by retention of Pol II in K562 As-treated cells. Boxplots showing Pol II reads at promoter of common DE genes, up genes and down genes at NHS state (left) and with As (right).

Contrary to this observation, in MCF7 cells these same genes behave differently. With HS, upregulated genes show increase in Pol II while downregulated genes show decrease in Pol II at promoter indicating decrease in Pol II recruitment at promoter. With As, this mechanism of Pol II retention at downregulated genes is similar to that of MCF7 cells (Fig 4.24).

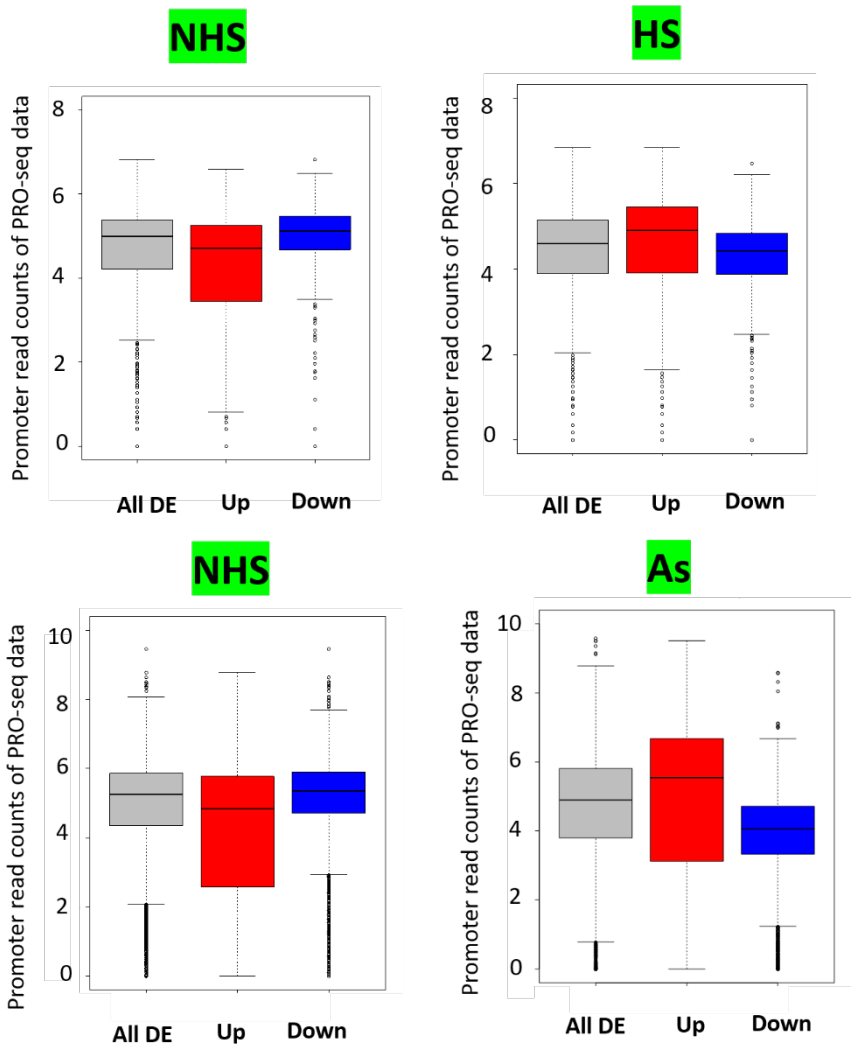


Figure 4. 24. Common downregulated genes between K562 and MCF7 show repression by decrease in recruitment of Pol II in MCF7 HS-treated (top) and As-treated (bottom) cells. Boxplots showing Pol II reads at promoter of common DE genes, up genes and down genes at NHS state (left) and with HS or As (right).

Hence, this result shows that genes are regulated by the properties of the cells rather than their own property. Same genes that are downregulated with HS or As in both MCF7 and K562 show different mechanisms of repression, depending on the cells.

4.11 MCF7 Cells Switch on Self-Compensatory Mechanism to Maintain Transcription on Prolonged Heat Shock

Since MCF7 cells showed different mechanisms of transcriptional repression upon HS, we were curious if these events are true for other time points too. To address this question, we analyzed MCF7 HS at 12min and at 30min and plotted PRO-seq signal for the activated and repressed genes. Interestingly, we found at earlier time points MCF7 cells exhibit similar mechanism of repression as K562 cells. HS-activated genes in MCF7 at earlier time points show increased recruitment of Pol II and increased release to gene body and HS-repressed genes show retention of paused Pol II at the promoter. However, with increased exposure to HS we see a sharp decrease in paused Pol II population at promoters of repressed genes. These results indicate that cells switch their mechanisms of gene activation or repression depending on the duration of the stress response and based on the availability of factors responsible for transcription (Fig 4.25).

4.12 A Majority of Pol II Bound Sites are Present at Ground State in MCF7 Cells

We next sought to address if Pol II peaks are acquired as a response to HS or if they are present in unstressed condition. To address that we did ChIP-seq against Pol II in NHS, HS, and As treated MCF7 cells to characterize the status of Pol II in different treatment conditions. We did not find major changes in the number of genome-wide Pol II peaks called in NHS, HS and As conditions. However, a slight decrease in Pol II peaks was observed with HS and As (Fig 4.26).

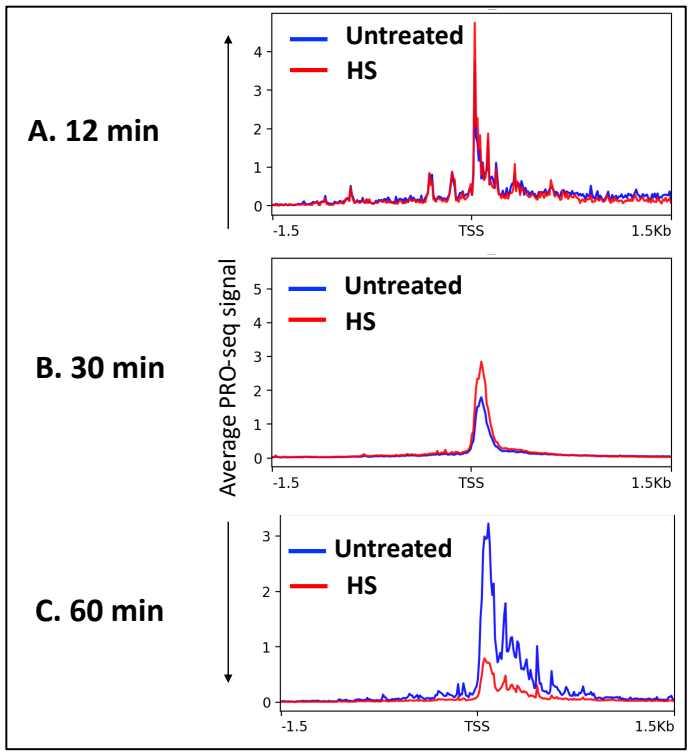


Figure 4. 25. MCF7 switch mechanisms of repression in response to HS in a time-dependent manner. PRO-seq profiles of repressed genes in
(A) 12 min of HS,
(B) 30 min of HS, and
(C) 60 min of HS showing dynamic changes in Pol II activity within the same cell responding to the same stress.

Moreover, overlapping Pol II peaks called in each condition points out the fact that most of the Pol II bound regions were already present at NHS state. 62% of the total number of peaks occurring in NHS, HS, and As combined are common among these three conditions indicating that genes are already present at a poised state ready for activation at ground state. With HS or As, 25% of the NHS Pol II peaks were lost (Fig 4.27).

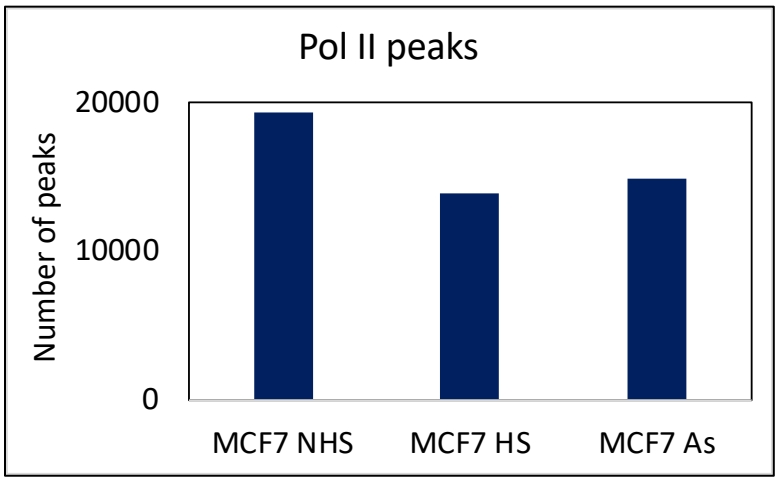


Figure 4. 26. Genome-wide Pol II peaks remain mostly unchanged with HS or As. Barplot representing the number of Pol II peaks called in each condition in MCF7 cells.

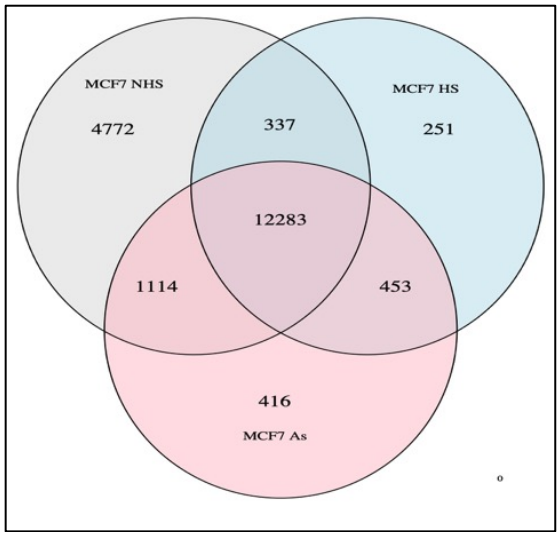


Figure 4. 27. Genome-wide Pol II binding sites are established at ground state. Venn diagram showing majority of the peaks overlap among the three conditions.

4.13 Dynamic Changes in Pol II Peaks at Promoter-distal Regions on Exposure to Heat and Arsenic

We next addressed where these Pol II peaks are located in the genome. We first annotated peaks that are called in each condition namely NHS, HS, and As separately and found, as was expected, majority of the peaks in each condition are located at the promoter regions (Fig 4.28).

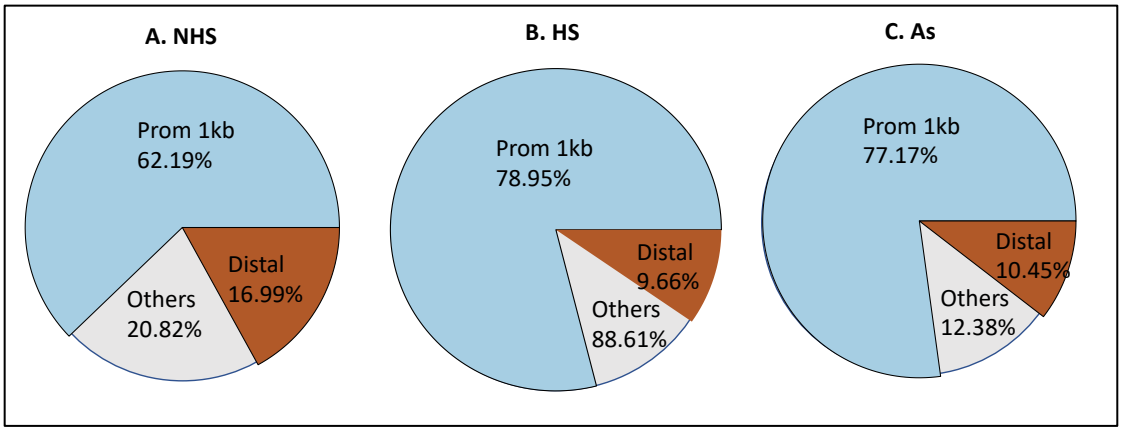


Figure 4. 28. Majority of Pol II peaks are located at promoter regions in all conditions. Venn diagram showing location of Pol II peaks in NHS (A), HS (B) and As (C) condition.

We then annotated the peaks that are exclusively present in NHS condition but are lost with either of the stresses. Surprisingly we found more than 54% of these peaks belong to promoter-distal regions and only ~ 14% of these peaks belong to the promoter (1kb) region (Fig 4.29).

We then annotated peaks that are present in all three conditions and found that 84% of these overlapping peaks belong to promoter regions (Fig 4.30).

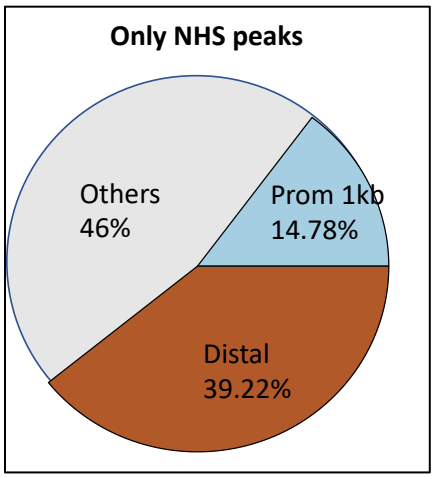


Figure 4. 29. Changes in global Pol II binding are more pronounced at promoter-distal regions. Venn diagram of unique Pol II peaks in NHS condition showing almost 3 times more enrichment of these peaks at promoter-distal regions than at promoter regions.

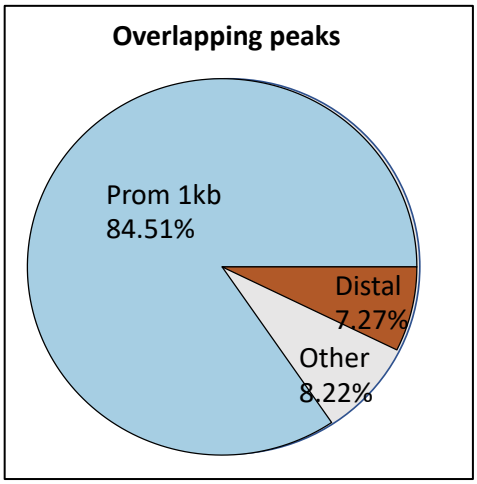


Figure 4. 30. Pol II peaks that are consistently present among all conditions belong to promoters. Venn diagram of overlapping Pol II peaks in NHS, HS, and As conditions showing almost significant promoter enrichment.

These results indicate that the distal intergenic regions are more dynamic in responding to stresses - 'potential enhancers' that do not get activated lose Pol II with stress and probably these Pol II get recycled and are recruited at promoters and enhancers of genes that are activated with stress.

4.14 Chromatin Openness and Pol II Occupancy are Correlated and Not Affected by Changes in Pol II Binding with Stress

We next sought to address if chromatin openness is associated with Pol II occupancy, as Pol II binding is known to be not very specific. We plotted Spearman correlation between Pol II reads at promoter (+/- 1kb) from ChIP-seq data, and ATAC-seq reads at gene promoter (+/- 1kb). Comparison between Pol II reads and ATAC-seq reads of all genes in hg19 human genome assembly database showed a high Spearman correlation of 0.79 (Fig 4.31 A, top). With HS, this correlation remains almost unchanged with a co-efficient of 0.72 (Fig 4.31 A, bottom). We next looked at genes that are either upregulated (Fig 4.31 B) or downregulated (Fig 4.31 C) with HS. Both categories of genes are well correlated in their Pol II occupancy and chromatin openness in NHS (Fig 4.31 top) and even in HS (Fig 4.31 bottom) state.

Pol II binding and ATAC-seq reads in As- treated MCF7 cells are also highly correlated between each other in both untreated (Fig 4.32 top) and As- treated conditions (Fig 4.32 bottom).

These results indicate that chromatin accessibility may influence Pol II binding, is stable and remain mostly unaffected with rapid stress responses.

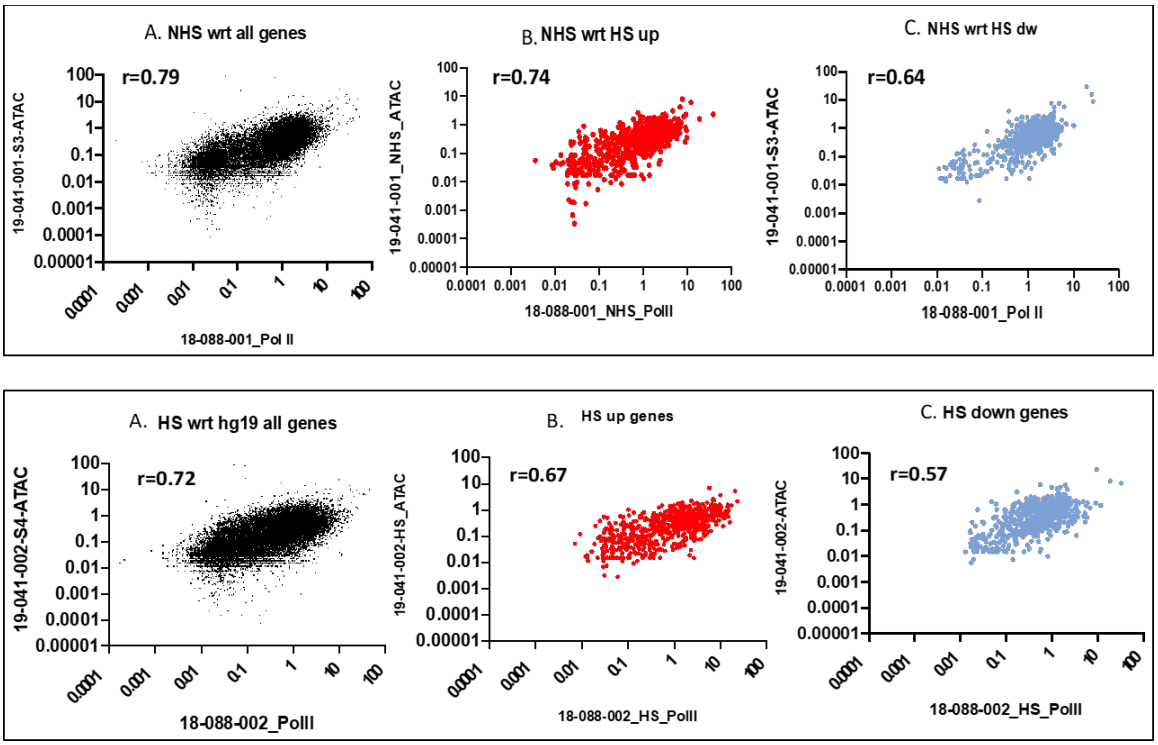


Figure 4. 31. ATAC-seq and Pol II peaks are highly correlated in NHS and HS treated MCF7 cells. Scatter plot distribution expressed in log-scale of all genes (A), HS-upregulated gene (B), and HS-downregulated genes (C) in NHS (top panel) and HS (bottom panel) conditions.

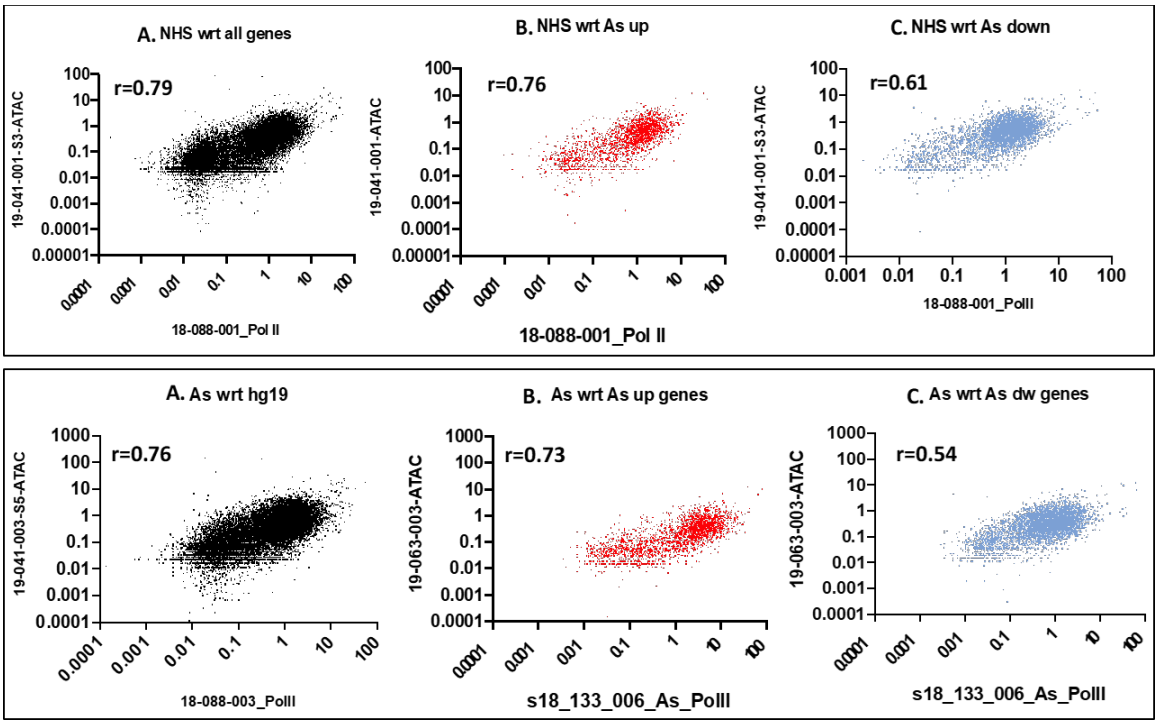


Figure 4. 32. ATAC-seq and Pol II peaks are highly correlated in NHS and As treated MCF7 cells. Scatter plot distribution expressed in log-scale of all genes (A), As-upregulated gene (B), and As-downregulated genes (C) in NHS (top panel) and As (bottom panel) conditions.

CHAPTER 5

IL-1β INDUCES GLOBAL TRANSCRIPTOMIC CHANGES IN DISTINCT ENHANCERS AND ASSOCIATED GENES IN A TIME-DEPENDENT MANNER

5.1 Transcription Regulation during Inflammation

Promoters and enhancers regulate transcription when cells are exposed to intracellular stimuli such as inflammatory molecules. Inflammation is a double-edged sword – inflammation needs to be precisely regulated in the cells to avoid continuous production of inflammatory molecules in the cells that can lead to various diseases including cancer. Hence it is critical to understand regulation of transcription in cells exposed to inflammatory molecules.

Lam *et al*., 2013 showed that nuclear receptors Rev-Erb-α and Rev-Erb-β that also act as transcriptional repressors can bind to enhancer regions in mouse macrophages and regulate eRNA synthesis. They further validated that gain or loss of these transcriptional repressors can lead to increased or decreased eRNA synthesis that cause differential expression of the target mRNA. However, this changes in binding pattern had no significant effect on histone marks further emphasizing the point that histones are stable safe keepers of chromatin and majority of them are not altered in rapid response reactions. In 2015, Hah *et al*. studied transcriptional responses in mouse bone marrow cells treated with LPS and identified specific super-enhancers that respond to LPS. They showed that eRNA transcription is dynamic and transcription is repressed when exposed to LPS for a prolonged period of time.

To study how inflammation affects transcription, we exposed human epithelial adherent non-small cell lung (NSCLC) adenocarcinoma cells A549 to proinflammatory cytokine IL-1β and performed RT-qPCR, Western Blots, RNA-sequencing, and PRO-sequencing.

5.2 Physiological Relevance of the Study

Inflammation plays a vital role in the development and progression of non-small cell lung cancer (NSCLC)[114,115,116]. It has been shown that infections and inflammation are associated with 15-20% of all cancer deaths[117,118]. A number of different types of immune cells such as natural killer cells (NK cells), T-lymphocytes, macrophages, dendritic cells (DC), myeloid-derived suppressor cells (MDSCs), and B-cells, infiltrate the tumor and along with stromal cells create a lung cancer microenvironment[119,120]. The chemokines and cytokines secreted by these tumor infiltrating cells determine the nature and progression of the respective tumors[120]. Pro-inflammatory cytokines such as IL-1α, IL-1β , IL-6, TNFα, TGF-β among multiple others are secreted by these immune cells and also by the tumor cells themselves and are released in the microenvironment[121]. These cytokines play a major role in cell proliferation, cell survival and metastasis of tumors[115].

Inflammation in lungs is not only contained within the organ, but it affects other organs as well. Recent studies suggest that lung inflammation has considerable contribution to heart diseases[122,123,124]. For example, chronic lung inflammatory response in rabbits directly corresponds to the volume of atherosclerotic plaque deposited on both left and right coronary arteries of heart[124]. Myocardial infarction in humans is nearly 5-fold more likely to be triggered by lung inflammation than inflammation in other organs such as acute urinary tract infection[125,126]. Studies show that there is a 30% reduction of

hospitalization and death from cardiovascular events in patients that had taken the influenza vaccine[127,128]. There are hypotheses indicating that acute lung inflammation leads to systemic inflammation causing release of cytokines such as IL-1β and IL-6 into the blood stream[122]. These pro-inflammatory cytokines cause destabilization of atherosclerotic plaques leading to cardiovascular events[129]. This evidence strongly indicates that cytokines released in lung inflammation plays a major role not only in lungs but also in promoting cardiovascular diseases. However, the mechanisms by which these cytokines regulate gene expression remain largely unknown.

It is, therefore, critical to dissect out the basic regulatory mechanisms of how genes and gene-regulatory regions respond to inflammatory molecules and bring about changes in global transcriptional landscape.

5.3 A549 Cells Mount up an Inflammatory Response when Challenged with IL-1β (Interleukin 1-Beta)

To identify which cytokines that are upregulated in lung cancers could mediate a robust proinflammatory stimulus in A549 lung epithelial adenocarcinoma cells, we treated these cells with various cytokines including IL-1β (Fig 5.1 A), IL-6 (Fig 5.1 B), and TNFα (Fig 5.1 C). We also used bacterial Pseudomonas lipopolysaccharide (LPS) (Fig 5.1 D) to stimulate pro-inflammatory response in A549 cells and measured mRNA changes using pre-reported inflammatory genes that have been shown to respond to these stimuli in other cell types.

Western Blot was performed on LPS-treated and IL-1β treated A549 cells to analyze the effect of these inflammatory molecules at the protein level. The two bands of LYN protein correspond to the two isoforms of the protein, 53KD and 56KD, respectively.

It is clear that when treated with inflammatory stimuli such as LPS (Fig 5.2 A) or IL-1β (Fig 5.2 B), Lyn and NFκB protein levels increase with time from 0 minutes to 5 hours. Phosphorylation of LYN is prominent with LPS at 30 min to 5h and with IL-1β at 5h (Fig 5.2).

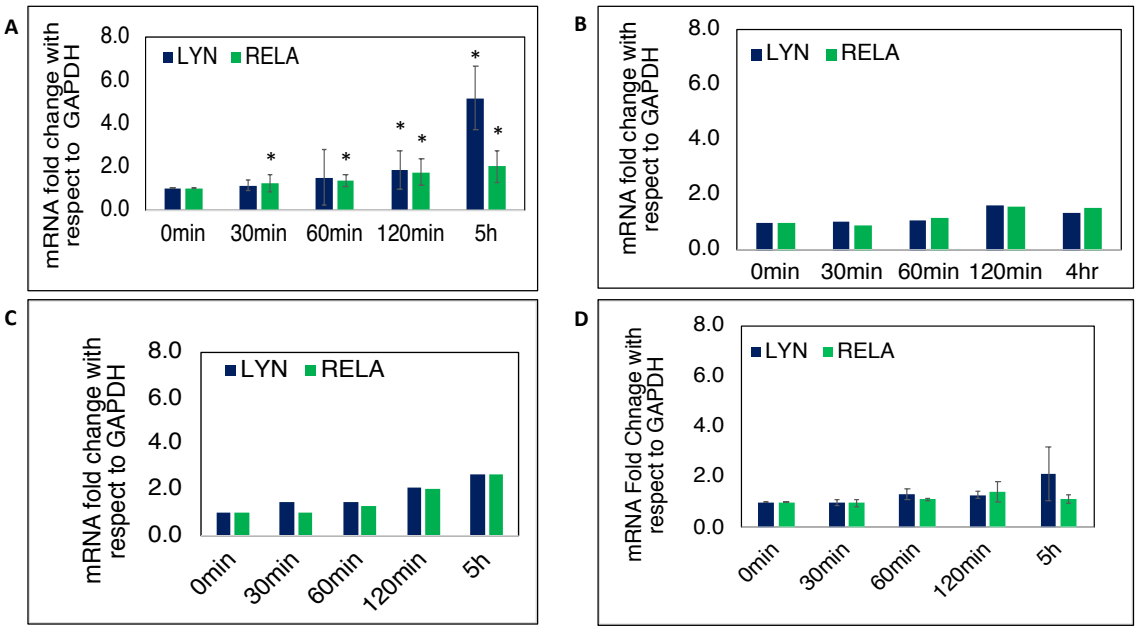


Figure 5. 1. A549 cells treated with a panel of inflammatory cytokines. RT-qPCR data showing LYN and RELA mRNA expression in A549 cells challenged in IL-1β (10ng/ml) (n=7) (A), IL-6 (10ng/ml) (n=1) (10ng/ml) (B), TNFα (10ng/ml)(n=1) (C), and LPS (100ng/ml) (n=2) (D). All data are analyzed using parametric paired t-test with a p-value cutoff of 0.05. Error bars represent standard deviation.

Since IL-1β mounts up a robust inflammatory response both at RNA and protein level in A549 cells among the cytokines tested, we conducted further experiments using A549 cells treated with IL-1β.

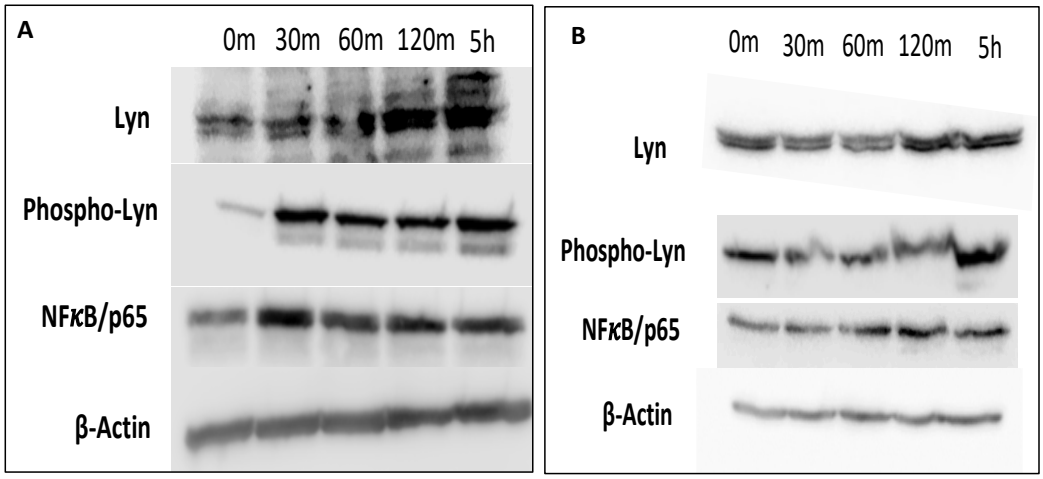


Figure 5. 2. Inflammatory molecules induce Lyn and NFκB protein levels in A549 cells. Western Blot in A549 cells induced with LPS (100ng/ml) (A) and IL-1ß (10ng/ml) (B) against Lyn, phospho-Lyn, and NFκB showing increased levels of these proteins with inflammation.

5.4 Validation of Duration for IL-1β Treatment

To address if longer exposure could lead to higher fold of gene activation, A549 cells were treated with IL-1β for a longer period of time (1-day, 3-day and 5-day). Protein and RNA were extracted for analysis (Fig 5.3).

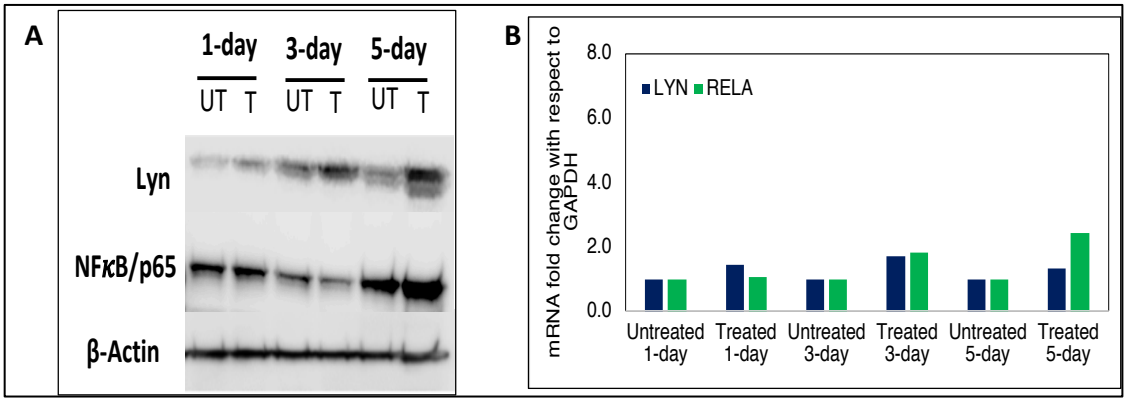


Figure 5. 3. Longer exposure to IL-1ß do not increase inflammatory response of A549 cells. Western blot (A) and RT-qPCR (B) on IL-1ß (25ng/ml) treated A549 cells (n=1) showing no significant change in Lyn or NFκB protein or mRNA level after 1-day or 3-day treatment.

Interestingly, from our western blot results, we observed that the level of inflammatory protein expression increases with the number of days the cells were cultured in tissue culture dishes even without any IL-1β treatment. Increase in LYN and NFκB protein and mRNA expression in 3-day or 5-day treatment was comparable to 1-day treatment.

Since our goal was to analyze how enhancers control promoter activation in response to IL-1β and we observed gene activation with 1-day treatment with no major changes in 3-day or 5-day treated cells, we selected the shorter time course.

5.5 Dose Response Curve of IL-1β Treatment

In order to have a concentration of IL-1β that is close to physiological concentrations of cytokines, the minimum concentration of IL-1β required by A549 cells to trigger inflammation was determined. A549 cells were exposed to 5 different concentrations of IL-1β: 0ng/ml (control), 5ng/ml, 10ng/ml, 25ng/ml and 50ng/ml (Fig 5.4). Cells were collected 3 days after exposure to IL-1β and mRNA and protein concentrations were measured against known target inflammatory genes.

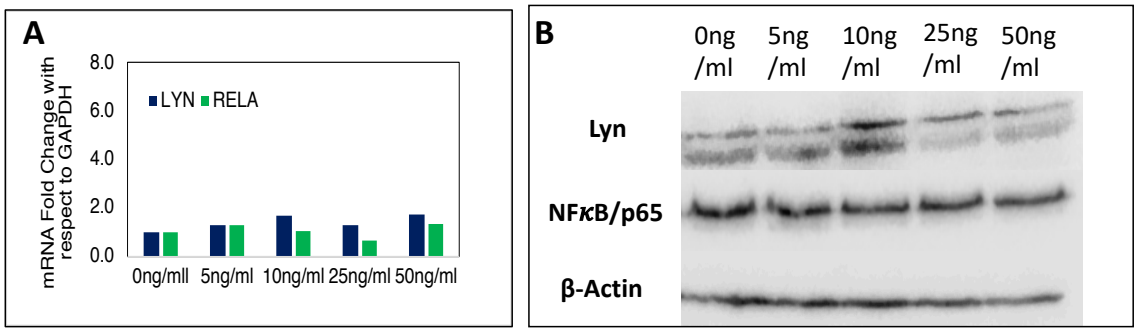


Figure 5. 4. Dose determination of IL-1ß to induce inflammation in A549 cells. RT-qPCR (A) and Western blot (B) on A549 cells (n=1) treated with a range of IL-1ß concentrations showing mRNA and protein upregulation at 10ng/ml.

From RT-qPCR and Western Blot analyses, we chose 10ng/ml as the minimum concentration of IL-1β to induce inflammation and have used this concentration for further experiments.

5.6 Role of IL-1β in Lung Inflammation and Cancer

The interleukin -1 or IL-1 is a family of multipotent cytokines that plays critical roles in physiological and pathological disease states in lungs and other organs of the body[130]. IL-1 family consists of 11 members among which two agonists, IL-1α and IL-1β, and an antagonist IL-1 receptor antagonist (IL-1ra), are well studied[131]. These cytokines are recognized by two major receptors, IL-1 receptor Type I (IL-1RI) and IL-1 receptor Type II (IL-1RII)[130,131,132]. IL-1RI is an 80-kD protein found on surface of all IL-1 responsive cells and is mainly involved in IL-1b signal transduction[130]. IL-1RII is a 60 kD protein bound to the cell membrane and like IL-1R1, shows greater affinity for IL-1 binding [130,132,133]. However, IL-1RII can undergo post-translational cleavage by metalloproteases, losing its cytosolic signal-transducing component, thus acting as a soluble form of inhibitor protein for IL-1β [130,131,132,133]. Therefore, IL-1RII acts as a decoy-receptor for regulating IL-1β level in cells and maintaining controlled inflammation.

The fine regulation of IL-1β level mediated by IL-1 receptors indicates that IL-1β has a major role in regulating inflammation and immune responses. Blockage or excessive synthesis of IL-1β has been shown to affect multiple autoimmune diseases such as Type-II Diabetes mellitus, Sjogren's Syndrome, Gout and Chondrocalcinosis, pericarditis, rheumatoid arthritis, chronic interstitial lung diseases, among many others[134].

The role of IL-1β in lungs is complicated. Human lung airway epithelial cells can respond to very low levels of IL-1β[132]. In some studies, bronchial epithelial cells have also been shown to produce IL-1β[133]. (This implies IL-1R I and IL-1RAcP (IL-1R accessory protein) act together to amplify the response in lung epithelial cells[133].) Human lung epithelial cells can express only IL-1 type I receptors and IL-1R antagonist and not IL-1 type II receptors. Therefore, the ability of these cells to regulate IL-1β activity is compromised and is dependent on inflammatory cells present in the milieu to downregulate IL-1β expression through release of IL-1β inhibitors[133].

These observations indicate that if the inflammatory milieu surrounding the bronchial epithelial cells are dysregulated and do not produce IL-1β inhibitors, the broncho-epithelial cells themselves will not be able to control IL-1β activity. This makes lung epithelial cells an especially interesting model system to study the rapid response to IL-1β. It is essential to characterize how this inflammatory cytokine regulates promoter activation and gene expression to understand the basic biology of the cytokine and translate the mechanisms to disease systems.

5.7 RNA-Sequencing Reveals Upregulation of Several Inflammatory Pathways in IL-1β Treated A549 Cells

A549 cells were treated with IL-1β for 0 minutes and 5 hours and RNA were sequenced from those cells. The 5-hour time point was selected for RNA-seq using poly-A selection as it is a measure of polyadenylated (mature) RNAs especially mRNAs that take time to accumulate in the cell. Principal component analysis (PCA) and heatmap show that treated cells differ from the untreated and also replicates of each condition cluster together (Fig 5.5).

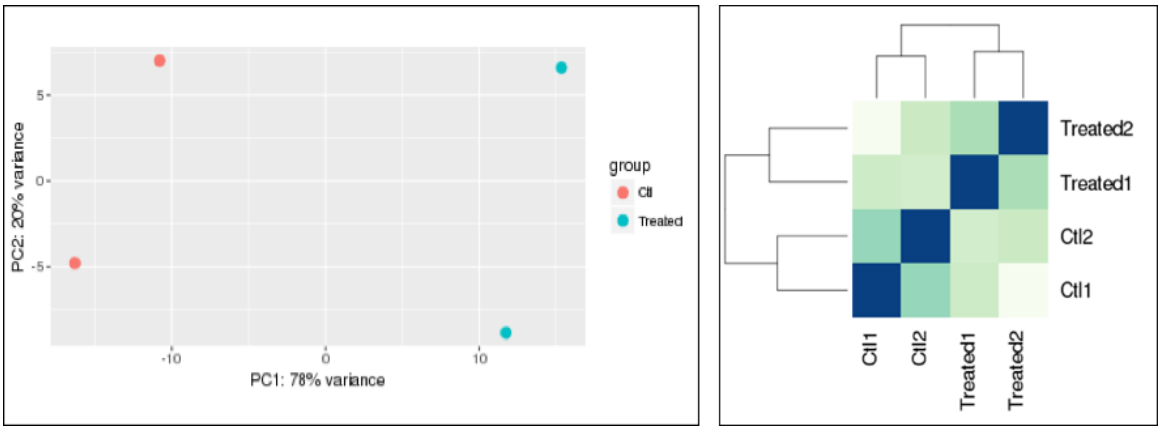


Figure 5. 5. Validation of replicates and treatments of RNA-seq samples. PCA plot (left) and heatmap (right) showing 5-hour IL-1ß (10ng/ml) treated A549 cells and the replicates cluster together than untreated A549 cells.

DESeq2 analysis was performed next on the two replicates to find differentially changed genes in response to IL-1β. With a stringent p-value cut-off of 0.01, we found 401 genes that were upregulated and 170 genes that were downregulated (Fig 5.6).

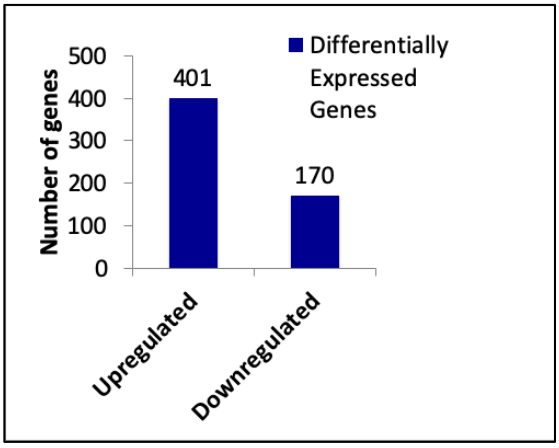


Figure 5. 6. DESeq2 analysis of IL-1ß treated A549 cells show robust changes in transcriptional landscape. Barplot showing the number of upregulated and downregulated genes in response to IL-1ß ($p < 0.01$).

Some genes with highest fold change for activation include IL-1α, IL-1β, and IL-6 among other interleukin family of genes. Our RNA-seq results validate the activation of signaling molecules involved in IL-1β signaling pathway in A549 cells, that has been previously reported in other cell types using various molecular techniques.

5.8 Signaling Mechanisms

IL-1 mediated signaling can occur in a MYD88-dependent canonical pathway[131] as well as a non-canonical MYD-88 independent pathway. In MYD88 dependent pathway, IL-1R1 binds IL-1 molecules and undergoes conformational changes in its extracellular domain. This structural change facilitates binding of IL-1RacP forming a trimeric complex that rapidly recruits myeloid differentiation primary response 88 (MYD88) and interleukin-1 receptor-activated protein kinase 4 (IRAK4). These five proteins together form a stable signaling complex. IRAK4 undergoes autophosphorylation and dissociates from the complex to form IRAK1 and IRAK2. Phosphorylated IRAK1 and IRAK2 recruit tumor-necrosis factor-associated factor (TRAF) 6 and together phosphorylates and activates downstream signaling molecules including TGF-β activated protein kinase 1 (TAK1) and MEKK3 signaling complexes. These downstream signaling complexes activate NF-kB, c-Jun N-terminal kinase (JNK) and p-38 MAPK pathways. In cytoplasm, p38 MAPK signaling molecules stabilize IL-1-responsive mRNA containing adenine-uridine-rich elements (AREs). This might be the reason we observe IL-1α and IL-1β stable RNA transcripts in our 5-hour treated A549 samples.

5.9 Pathway Analysis of Upregulated Genes using Reactome Pathway Database

KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis of these IL-1β upregulated genes show most of the upregulated genes fall into cytokine-cytokine receptor interaction pathway, TNF-signaling pathway, and NF-kappa B signaling pathway (Fig 5.7). This result further validates that IL-1β is a potent inducer of inflammation in A549 cells.

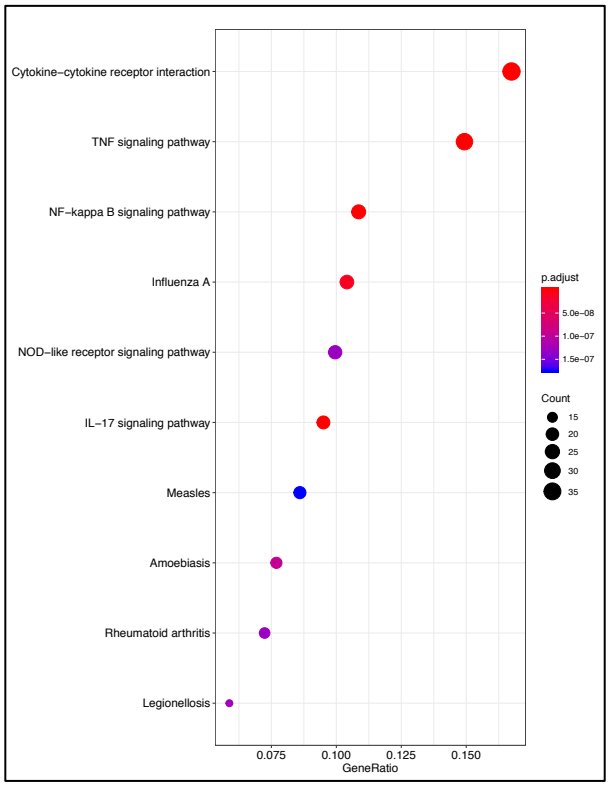


Figure 5. 7. KEGG pathway analysis of IL-1ß upregulated genes showing IL-1ß induces genes involved in several inflammatory pathways.

Among the upregulated genes, IRAK1 is an essential component of the canonical MYD-88 dependent pathway. However, upregulated genes also include several cytokines (as is seen in the KEGG analysis) including IL-18R and MAPK that might be involved signal transduction via MYD-88 independent pathway in these cells; but this prediction needs to be validated by further experiments. Hence, A549 lung cancer cells respond to IL-1β and activate various cytokine signaling pathways.

5.10 IL-1β Downregulates cAMP-Signaling Pathway in A549 Cells

Unlike in female granulosa-lutein cells of the human ovary where IL-1β upregulates the phosphorylation of cAMP-response element binding protein (CREB)[132], in lung cancer A549 cells IL-1β represses genes in cAMP pathway and Cushing's syndrome (Fig 5.8). This is a novel observation and needs further studies to uncover how cAMP signaling pathway is suppressed in A549 cells and what implication it might have in cell biology.

5.11 PRO-Seq Reveals a Disconnect between Promoter-Proximal and Gene Body Transcription in Response to IL-1β

Since RNA-seq is good only for detecting stable poly-A tailed RNAs, we were unable to detect immediate genome-wide transcriptional changes in A549 cells treated with IL-1β. In order to identify nascent transcriptomic changes, we performed PRO-seq on untreated A549 cells and cells treated with IL-1β for a short time period of 30 min and a long time period of 5h. Libraries were first sequenced in MiSeq and then in HiSeq, as is mentioned in Methods, and BAM and BIGWIG files were generated using the PRO-seq pipeline. BAM files were then processed using published NRSA (Nascent RNA Sequencing Analysis) pipeline (Wang et al., 2018) for identification of nascent

transcription. Reads were normalized using sequencing depth and a FDR cut off of less than 0.05 was set.

A gene is divided into two parts – promoter-proximal region (pp) which is defined as TSS +/- 500bp and gene body (gb) regions which is defined as TSS +1kb to TTS (transcription termination site) (Fig 5.9).

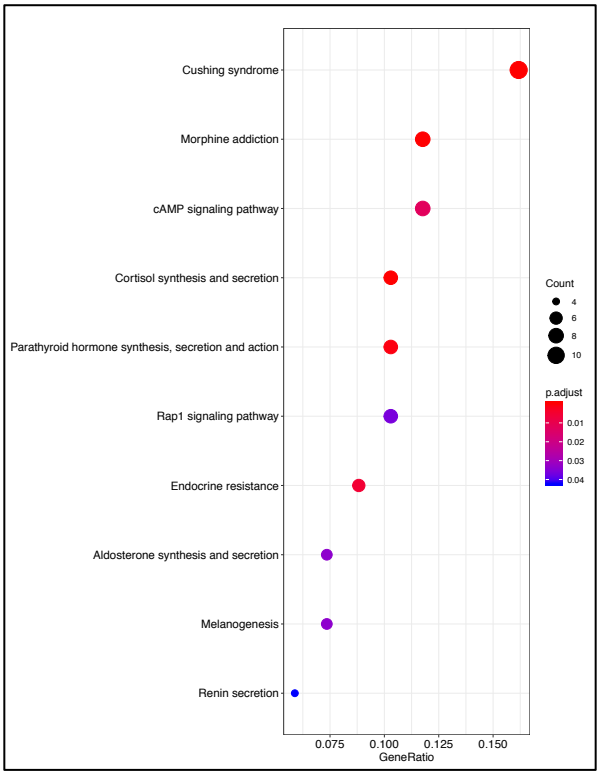


Figure 5. 8. KEGG pathway analysis of IL-1ß downregulated genes showing IL-1β represses genes involved calcium signaling pathway in A549 cells.

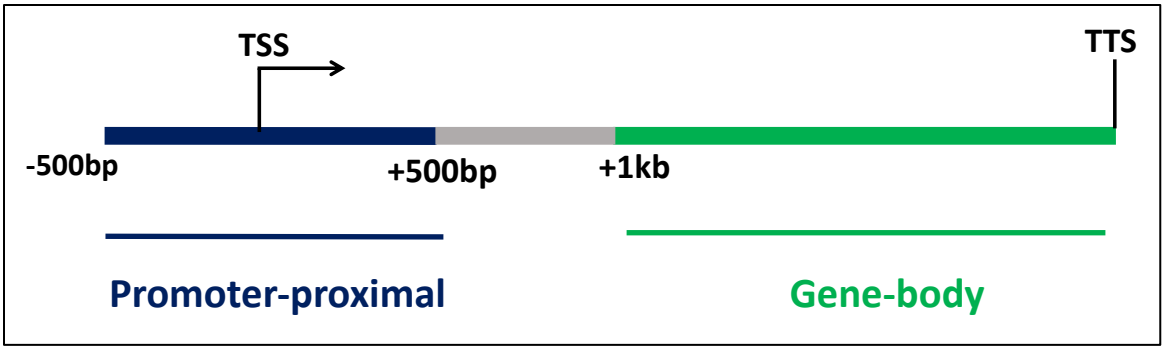


Figure 5. 9. Schematic representation of a gene divided into promoter and gene body, as has been used for PRO-seq analysis.

Analysis of reads at promoter-proximal region versus reads at gene-body region reveal clear discrepancy between promoter-proximal transcription and gene-body transcription in IL-1β treated cells. At 30 min we found changes in 1880 promoter-proximal regions but only 758 gene body regions as compared to 0 min. Similarly, at 5h we found differential expression at 2049 promoter-proximal regions but 763 gene body regions (Fig 5.10).

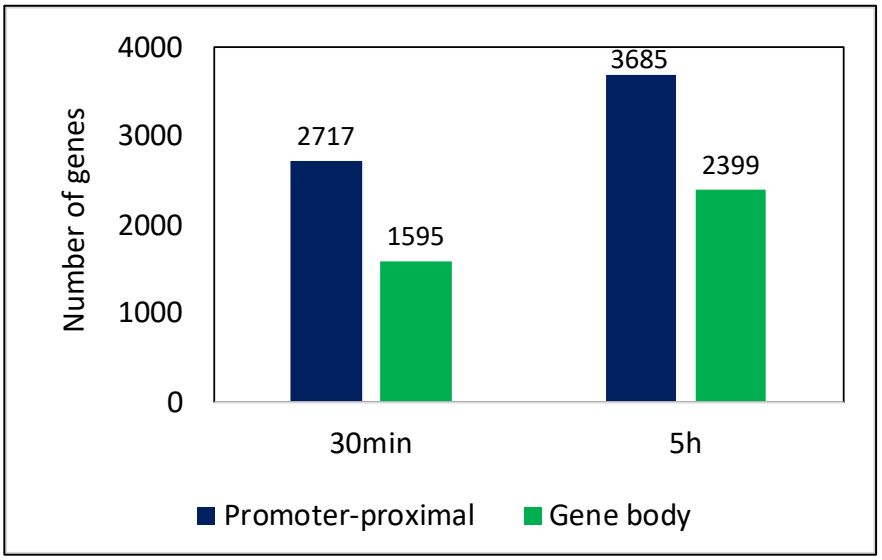


Figure 5. 10. Evidence of abortive transcription at promoter-proximal regions of gene. Barplot showing promoter proximal transcription is higher than gene body transcription at both time points.

This finding points to the phenomenon of premature termination and abortive transcription that has been previously in *Drosophila*, HIV and other eukaryotic systems. In response to IL-1β, transcription initiates at several promoters. However, most of these transcriptions are aborted and only genes that are specific to respond to IL-1β ultimately undergo full transcription. How this specificity is determined or if these premature transcripts have any specific functions are still open questions and would need further experiments.

5.12 Majority of the Genes Body Reads Overlap with the Promoter-Proximal Transcripts

We next asked how many of these genes that start transcription produce a full-length transcript. To address that, we overlapped promoter-proximal reads with gene body reads and found 52% and 68% of the gene body transcripts at 30 min and 5h respectively are extensions of the promoter proximal reads (Fig 5.11).

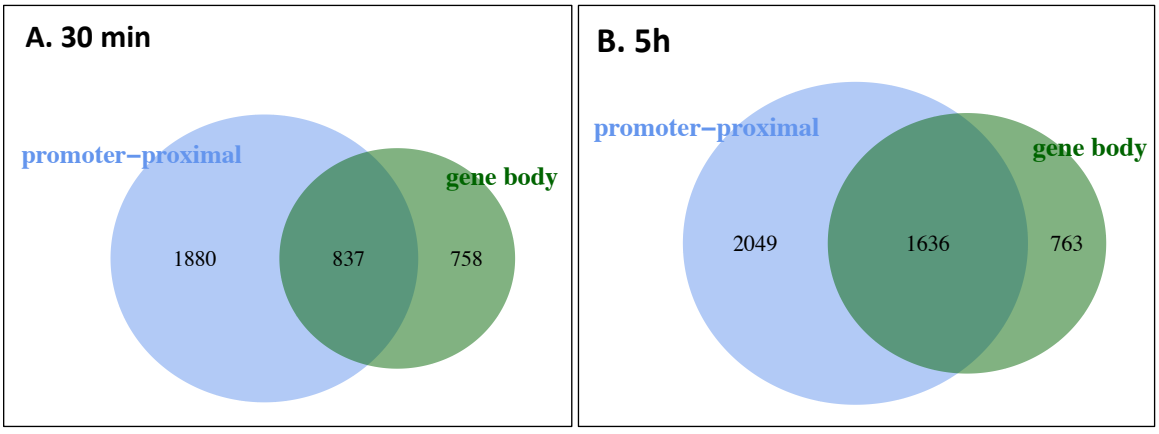


Figure 5. 11. Majority of IL-1β induced transcripts are overlapping between promoter and gene body. Venn diagram showing overlap between promoter-proximal transcript and gene body transcript at 30 min (A) and 5h (B).

5.13 Classification of Genes Based on Patterns of Transcription

Based on promoter-proximal versus gene body transcription, IL-1β responsive genes can be clearly categorized into three distinct categories:

(a) Genes that show only promoter proximal transcription

(b) Genes that show only gene body transcription

(c) Genes that show both promoter proximal and gene body transcription

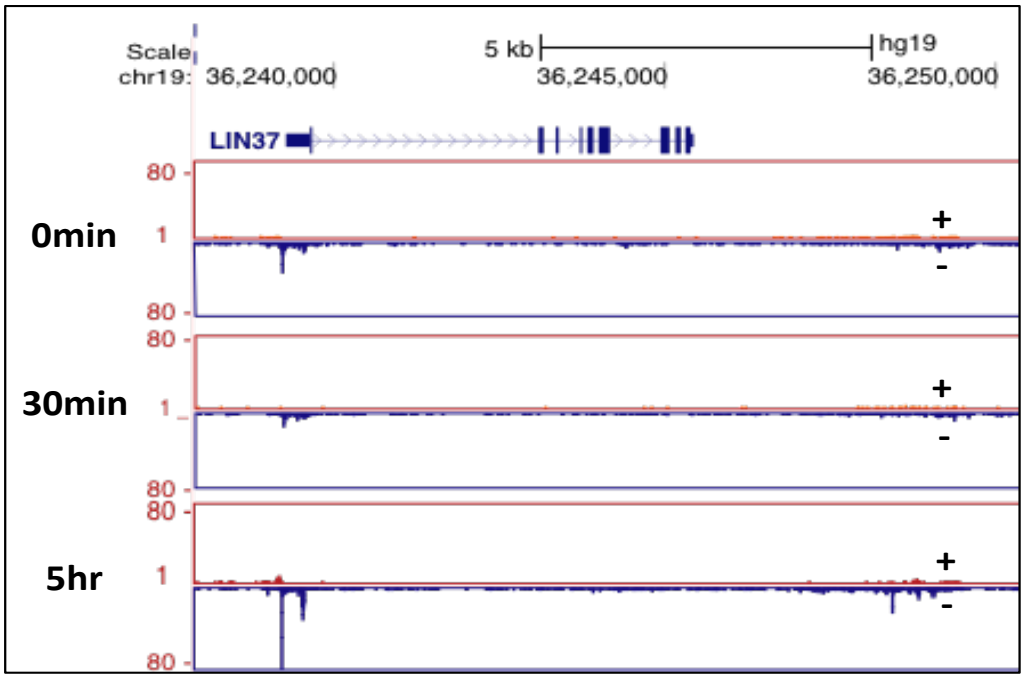
Below are UCSC browser tracks of an example gene from each category to highlight the differences in transcriptional readouts (Fig 5.12).

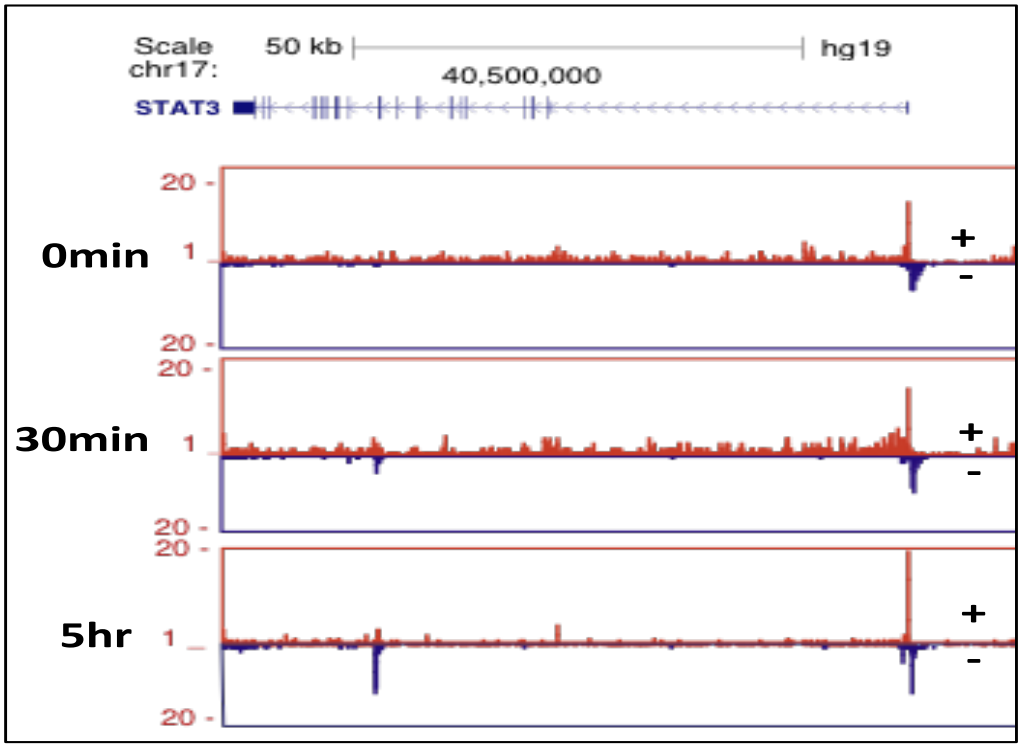
5.14 Gene Body Counts and Promoter-Proximal Counts among Different Replicates and Time-Points are Well Correlated

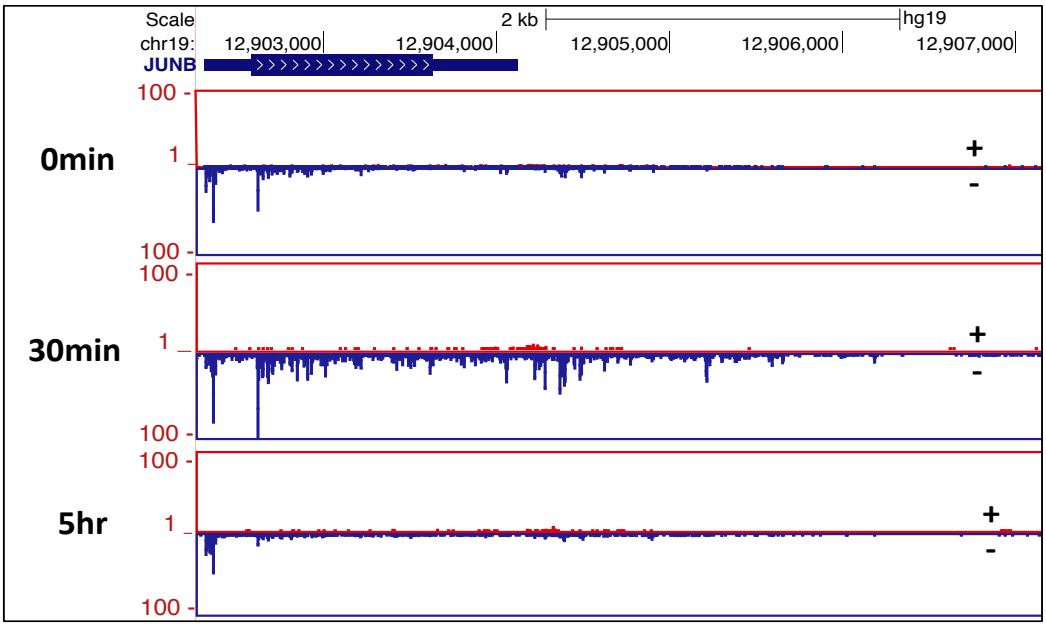
To validate that replicates are correlated with each other, we performed Spearman correlation test on gene body counts and correlation between replicates is between 80% and 90% and correlation between different time points is around or more than 70% (Fig 5.13)

We ran the same analysis for promoter-proximal reads and found correlation is consistent among different replicates and time-points of IL-1β treated A549 cells (Fig 5.14).

Figure 5. 12. Three classes of genes categorized based on transcriptional activity. UCSC browser shot of gene with only promoter proximal transcription (top), only gene-body transcription (middle), and both promoter proximal and gene body transcription (bottom).


Scale
chr19:
36,240,000
5 kb
36,245,000
hg19
36,250,000
LIN37
0min
30min
5hr
80
1
+
-


Scale
chr17:
50 kb
40,500,000
hg19
STAT3
0min
30min
5hr
20
1
+
-


Scale
2 kb
hg19
chr19:
12,903,000
12,904,000
12,905,000
12,906,000
12,907,000
JUNB
0min
30min
5hr
100
1
+
-

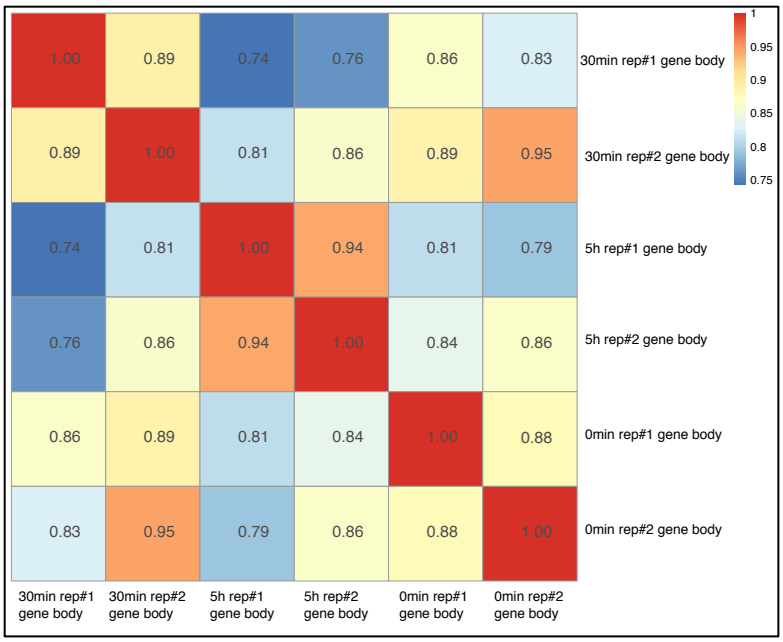


Figure 5. 13. IL-1β induced gene body transcription is well correlation between replicates and also among different time points. Heatmap with spearman correlation values in small squares between different replicates and different time-points showing strong correlation values.

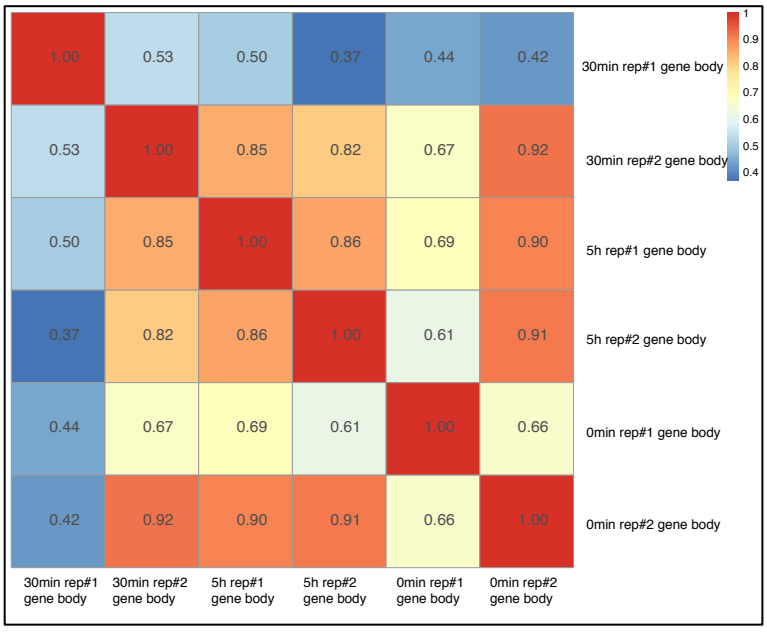


Figure 5. 14. IL-1β induced promoter proximal transcription is well correlation between replicates and also among different time points. Heatmap with spearman correlation values in small squares between different replicates and different time-points showing good correlation values.

5.15 Classification of Genes Based on Temporal Response to IL-1ß

Depending on the temporal activation or repression of transcription in response to IL-1β, genes can be classified as

a) Early response genes – genes that are differentially regulated at 30 min

b) Delayed genes – genes that are differentially regulated at 5h

5.15.1 Immediate Early Genes

Genes that have their gene body counts either increased or decreased with a fold-change cutoff of 1.5 and FDR < 0.05 at 30 min compared to 0 min treated IL-1β are characterized as immediate early genes. We find that there are 185 genes that are upregulated and 232 genes that are downregulated (Fig 5.15).

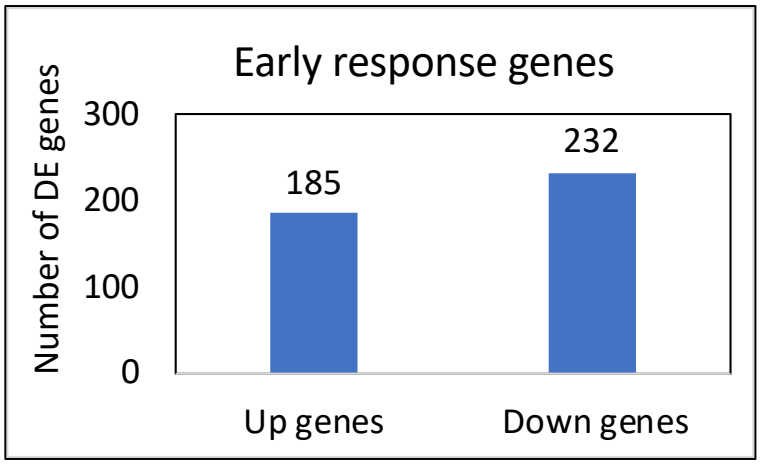


Figure 5. 15. IL-1β induces rapid changes in transcriptome inducing early genes including IEGs. Barplot showing numbers of activated and repressed genes within 30 min of IL-1β exposure.

Some of the early activated genes also include ‘immediate early response’ genes or IEGs such as JUNB. JUNB is known as a proto-oncogenic transcription factor and participates in IL-1β mediated signaling pathways. Hence this shows that IL-1β can trigger proto-oncogenic genes in A549 cells including activation of immediate early genes involved in IL-1β and NFκB signaling pathways.

Pathway analysis of genes that are upregulated at 30 min show increase in transcripts of genes involved TNF-α, IL-17, and NFκB signaling pathways. We conclude that IL-1β triggers classical inflammatory pathways in A549 cells as early as within 30 min of exposure (Fig 5.16).

5.15.2 Delayed Response Genes

PRO-seq analysis of gene body counts of differentially expressed genes at 5h reveal that there are 263 genes that are upregulated, and 154 genes are downregulated. (Fig 5.17).

Pathway analysis of upregulated genes show that most of the inflammatory genes do not show enrichment except TNF. However, delayed response genes such as genes involved in metabolic pathways are upregulated at 5h. These genes possibly get activated later to maintain homeostasis in stressed cells.

We next asked how many of the genes activated at 30 min are still active at 5h. We used R intersect function and found 43% of the genes activated at 30 min are still active at 5h (Fig 5.19).

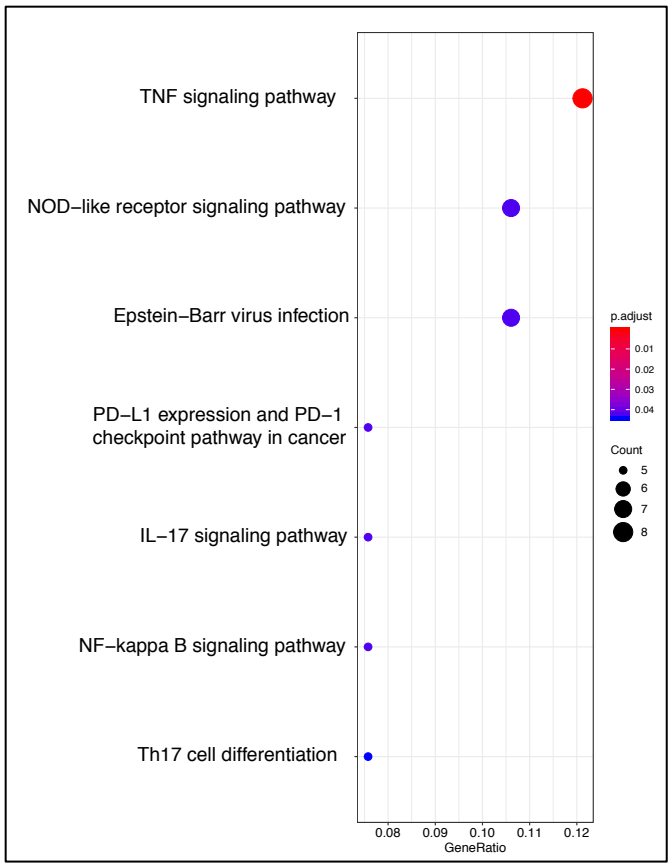


Figure 5. 16. KEGG pathway analysis showing early genes activated belong to inflammatory pathways.

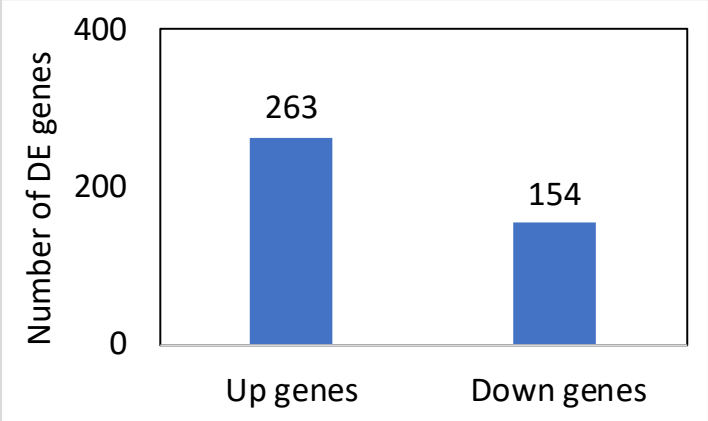


Figure 5. 17. PRO-seq reveals nascent changes in transcriptome at 5h of IL-1β induction. Barplot showing numbers of activated and repressed genes at 5h of IL-1β exposure.

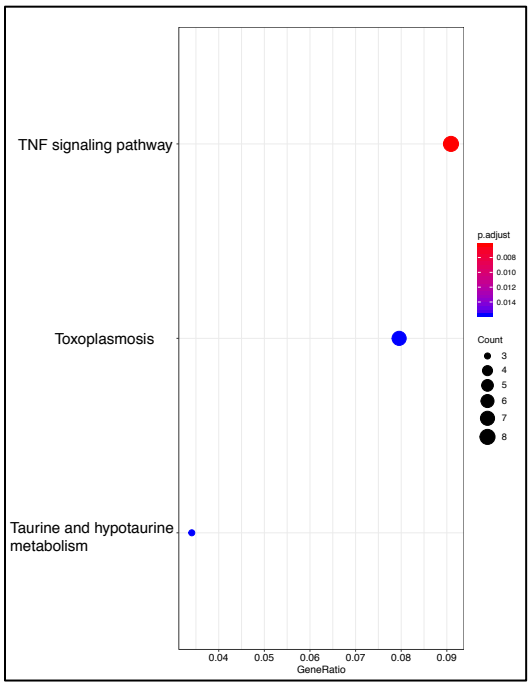


Figure 5. 18. KEGG pathway analysis showing genes activated at 5h are related to metabolism.

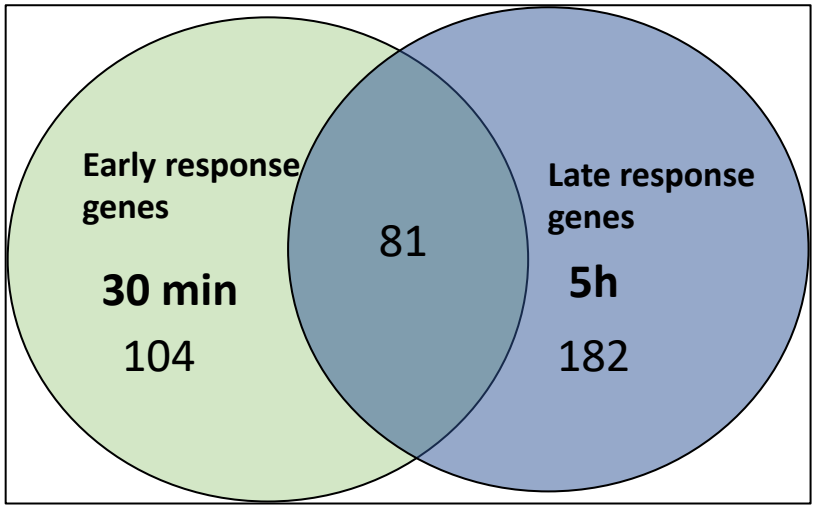


Figure 5. 19. Prolonged expression of few early activated genes. Venn diagram showing overlap between genes activated at 30 min and at 5h.

Although pathway analysis of all genes upregulated at 5h showed only genes involved in TNF-α pathway, analysis of only the 81 genes that are common between both time points showed consistent activation of the inflammatory pathways that were already activated at 30 min (Fig 5.20).

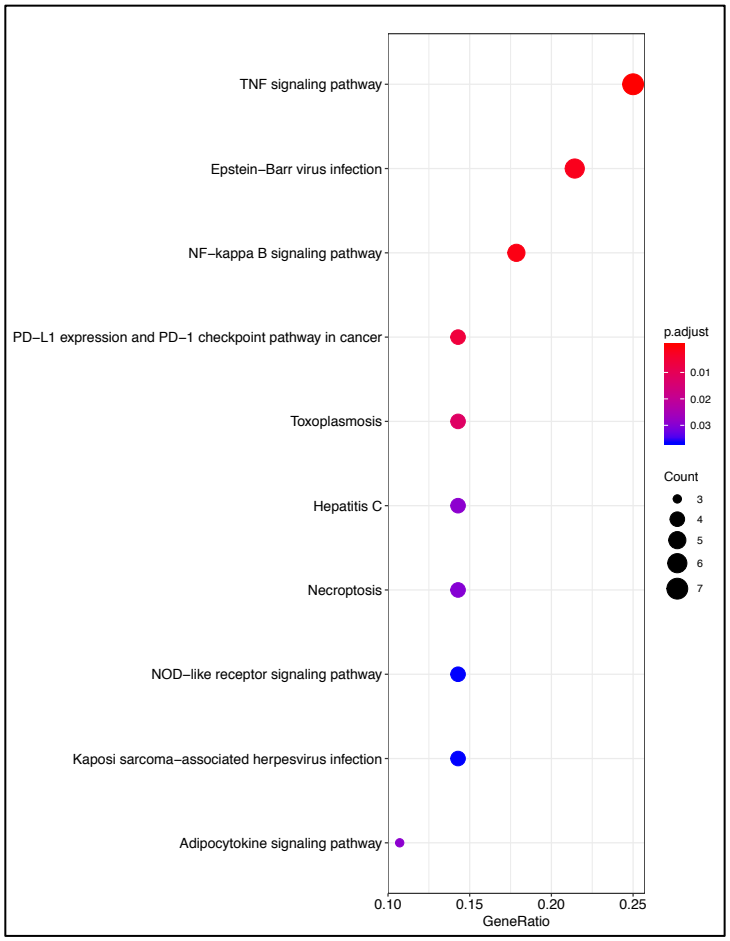


Figure 5. 20. Prolonged expressed genes belong to inflammatory pathways. KEGG pathway analysis showing active genes common between 30 min and 5h belong to several inflammatory and infection-related pathways.

5.16 Early Genes are Associated with High Enhancer RNA Transcription

Immediate early genes that are rapidly upregulated are known to have stronger promoters, shorter transcription units and are recently found to be associated with enhancers[135]. We were curious to see if any of these early IL-1ß activated genes by are associated with predicted enhancers or not. We chose only those genes that were closest to the enhancers (less than 50kb) as the gene-associated with enhancer. Interesting we found that 4 out of the 10 enhancers with the highest fold change in eRNA transcription are associated with gene body changes of early genes (Fig 5.21).

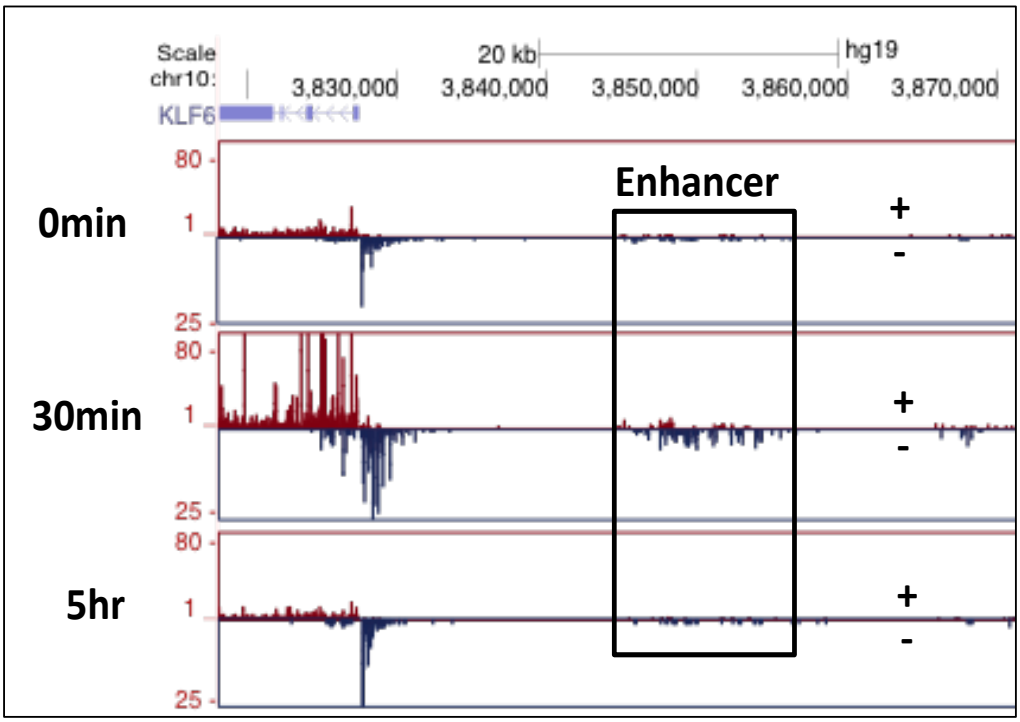


Figure 5. 21. Early activated gene KLF6 show increased transcription of enhancer region. UCSC browser shot of PRO-seq data of KLF6 gene with its enhancer region showing upregulation of mRNA and eRNA expression with IL-1β.

5.17 IL-1β Caused Rapid Changes in Enhancer RNA Transcription

Next, we looked for genome-wide changes in enhancer RNA transcription with IL-1β and found that 442 predicted enhancers are upregulated and 406 are downregulated with a fold-change cutoff of 1.5 within 30 min of IL-1β exposure. With 5h, the number of differentially expressed enhancer RNAs do not undergo dramatic increase or decrease meaning that enhancers are probably activated quite rapidly in response to stress or inflammatory stimuli (Fig 5.22).

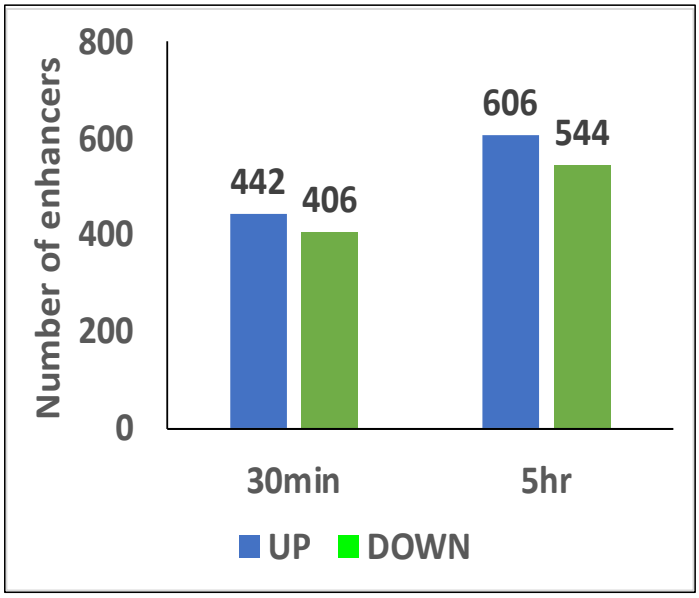


Figure 5. 22. IL-1β exposure leads to dynamics changes in eRNA transcription in A549 cells. Barplot showing differentially expressed enhancers at 30 min and 5h.

5.18 KLF6 eRNAs are Activated in Response to Inflammatory Cytokines Although the Timing of Activation Varies with Cytokines

We next asked whether eRNA transcription is specific only to IL-1β stimulus or is it upregulated with other cytokines in IL-1β pathway. We designed qPCR primers targeting predicted upregulated enhancers such as KLF6 and ARL8A (Fig 5.23).

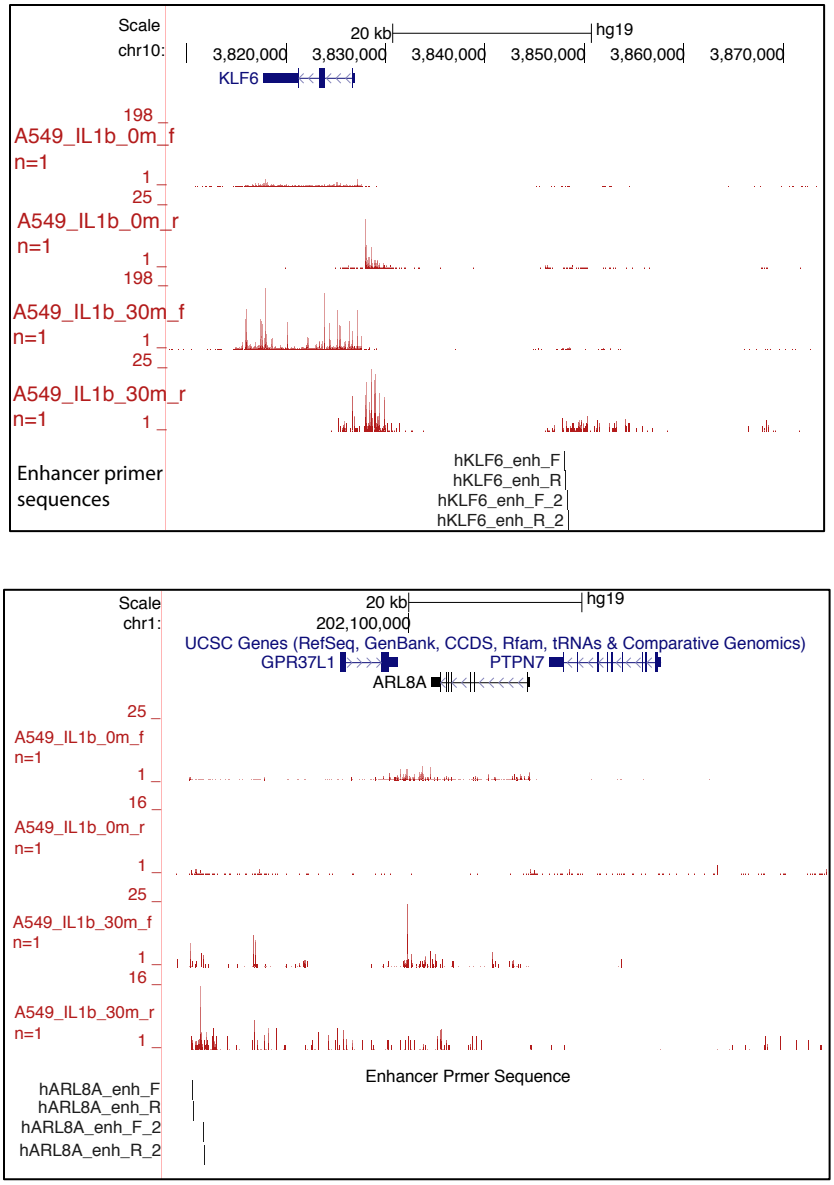


Figure 5. 23. Primer sequences for eRNA detection by qPCR. UCSC genome browser showing primers designed against enhancer regions for KLF6 (top) and ARL8A (bottom) that show upregulated expression with IL-1β.

We performed qPCR (n=1) in A549 treated with IL-1β in short intervals of time: 0 min, 30 min, 60 min, 120 min, and 5h. We detected upregulation of mRNAs of target inflammatory genes such as EGFR and RELA. We also detected upregulation of KLF6, however, we could not detect change in expression of ARL8A mRNA using qPCR. We

detected upregulation of KLF6 eRNA at 30 min, but a longer time-course led to decrease of KLF6 eRNA (Fig 5.24). Since we were not able to able to detect ARL8A mRNA change with qPCR, we focused on dynamics of KLF6 mRNA and eRNA. Furthermore, recently it has been confirmed by Syafruddin et al., 2019 using CRISPR deletion that our predicted KLF6 enhancer region is actually a super-enhancer that drives KLF6 gene expression in renal carcinoma cells[136].

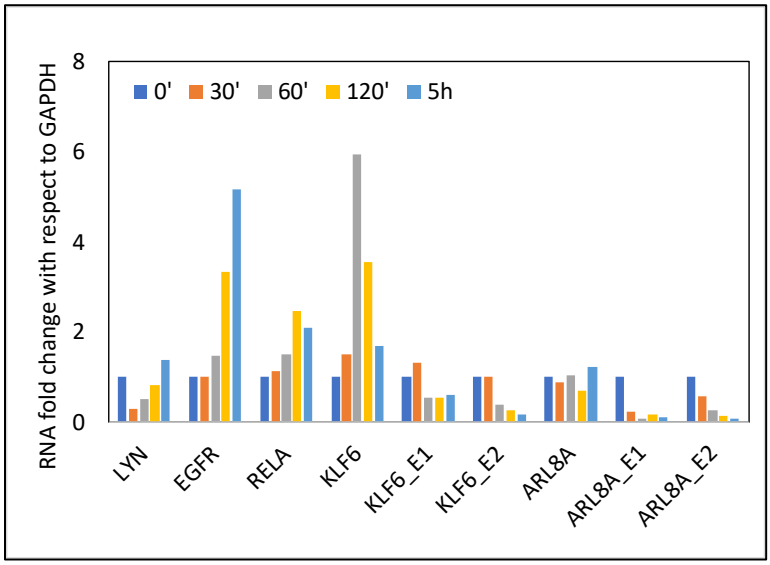


Figure 5. 24. KLF6 eRNA level goes down after 60 min. RT-qPCR data against several IL-1ß-target mRNAs and eRNAs showing eRNAs are not detectable/ goes down after 60min of IL-1ß-treatment in A549 cells.

We treated A549 cells with other cytokines such as IL-6 (n=1) and TNF-α (n=1) that were upregulated in response to IL-1β as detected by PRO-seq and RNA-seq. For both IL-6 and TNF-α, we found detectable KLF6 eRNA upregulation, although the timing of upregulation was much later than what we found in case of IL-1β (Fig 5.25).

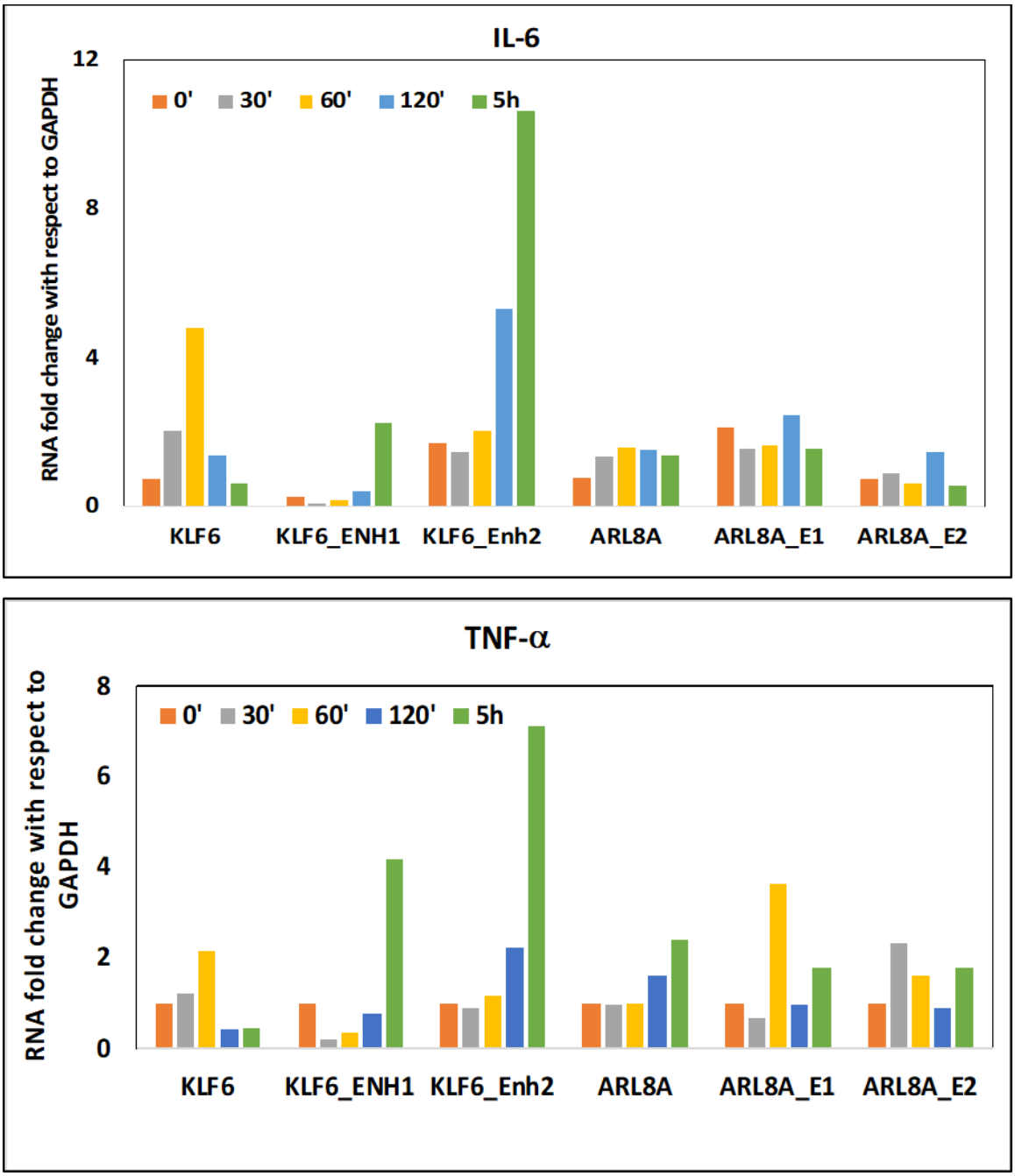


Figure 5. 25. Inflammatory molecules upregulated in IL-1ß pathway is able to induce KLF6 eRNA transcription. RT-qPCR results showing IL-6 (10ng/ml) (top) and TNF-α (10ng/ml) (bottom) induce KLF6 mRNA and eRNA expression in A549 cells.

5.19 KLF6 eRNA and KLF6 Pre-mRNA are Co-Transcribed and are Detected within 15 Minutes of IL-1ß Treatment

It is controversial in the field whether eRNA is expressed before mRNA is transcribed or simultaneously with mRNA and still can affect mRNA expression. From previous results we detected KLF6 eRNA upregulated at 30 min and we saw a burst of KLF6 mRNA at 60 min. We were curious if eRNA is actually synthesized prior to mRNA in response to IL-1β. To address that, we designed primers against unspliced, pre-mRNA state of KLF6 along with processed mRNA transcript. We also designed pre-mRNA primers against ACTB, a gene that does not respond to IL-1β as a negative control. Next, we treated A549 cells for even shorter time points - 0min, 15 min, 30min, 45 min, and 60 min and extracted RNA from them. We performed qPCR on these samples, and we detected changes in pre-mRNA transcription as early as in 15min, however the major detectable changes in pre-mRNA and eRNA was observed at 30 min and mRNA changes was pronounced at 45 min. As expected, we did not see any changes in the expression of negative control ACTB pre-mRNA (Fig 5.26).

Since detection of low abundance transient eRNA transcripts by qPCR is a method prone to artefacts unless properly controlled, we ran the qPCR end products with KLF6 eRNA primers on a 6% polyacrylamide gel with 100bp ladder. We found that the amplified products exactly match the length of the region amplified by the target primers indicating that these transcripts are specific (Fig 5.27).

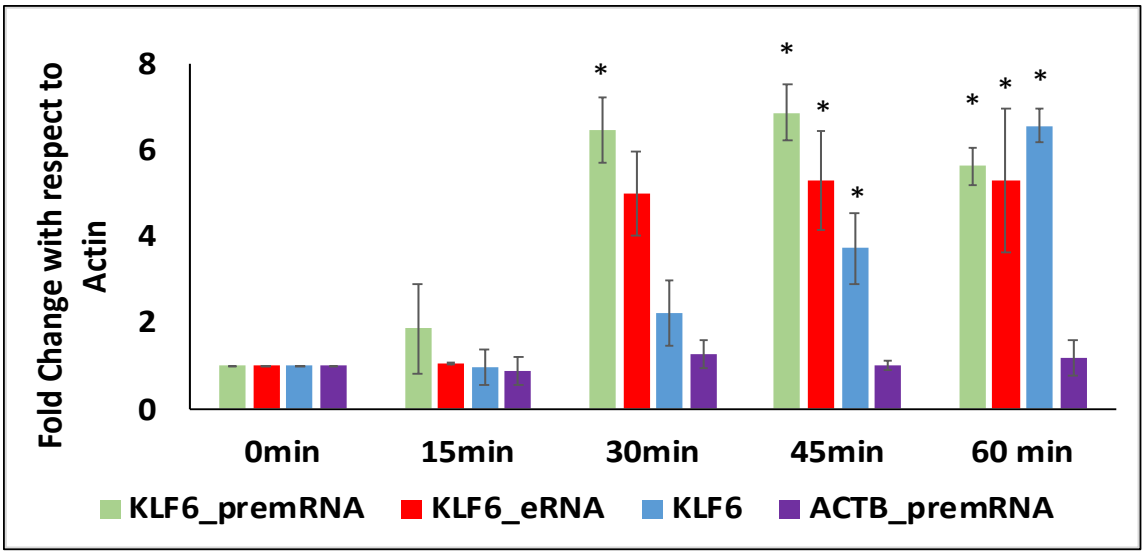


Figure 5. 26. KLF6 mRNA and eRNA are activated concomitantly with IL-1β induction. RT-qPCR results showing KLF6 pre-mRNA and KLF6 eRNA expression are both detectable at the same time without any lag in mRNA expression after eRNA activation.

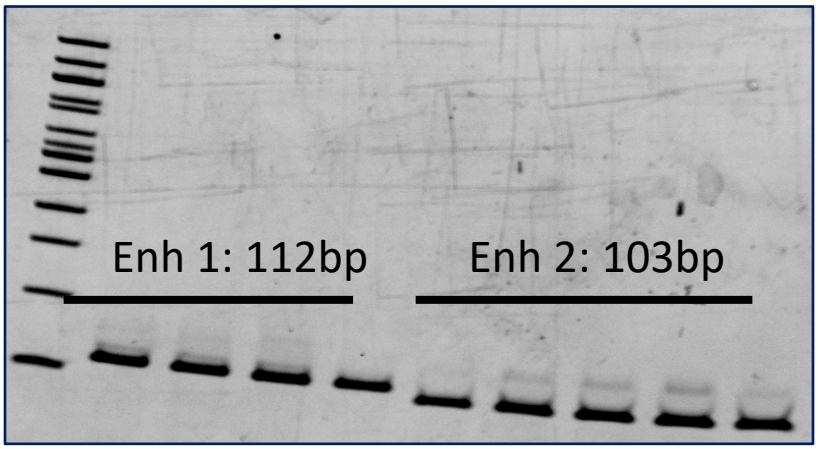


Figure 5. 27. KLF6 eRNA amplification is specific and matches expected amplicon size. KLF6 eRNA transcripts run on a 6% polyacrylamide gel showing expected sizes for the respective amplicons.

5.20 PRO-Sequencing at Shorter Time-Points Confirm Dynamic Changes in KLF6 eRNA Activation with IL-1β Treatment as Early as 15 Minutes

We further performed PRO-sequencing (n =2) of IL-1β treated A549 cell in a shorter time course of 0 min, 15 min, 30 min, 45 min, and 60 min (Fig 5.28). As expected, we detected a steady increase in nascent transcription of KLF6 eRNA from 15 min to 60 min of IL-1β treatment, as was previously detected by qPCR (Fig 5.26).

Hence, we conclude that -

(i) with IL-1β treatment in A549 changes we could see detectable time-dependent upregulation in KLF6 eRNA

(ii) eRNA and mRNA are co-transcribed and there is no gap in timing of transcription, as is detectable by the resources we used.

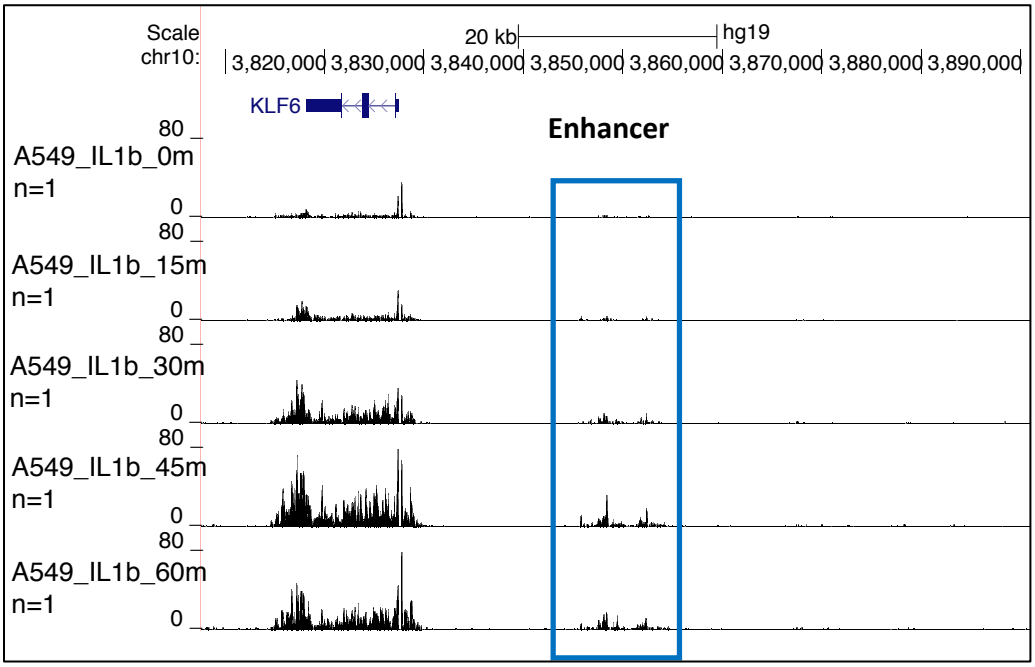


Figure 5. 28. KLF6 eRNA activation can be detected as early as after 15 min of IL-1β exposure in A549 cells. UCSC genome browser of PRO-seq data showing transcriptional changes at KLF6 enhancer region (blue box) and KLF6 gene at different time points of IL-1ß exposure.

CHAPTER 6

DISCUSSION

The primary focus of this study is to determine genome-wide changes in the human transcriptome during rapid-response to a stimulus. Before the era of next-generation sequencing, rapid-response studies were conducted looking at individual genes and their changes in expression. Consequently, changes outside the genes were neglected and intergenic regions were referred to as "junk DNA". Genome-wide sequencing studies led to a paradigm shift and the field started realizing that these "junk DNA" are, in reality, regions that regulate the genes.

In the last decade whole-genome mRNA sequencing evolved as the most popular method to visualize global changes in transcription in responses to stimuli. However, RNA-sequencing could not provide accurate measurements of the genome-wide nascent transcription that took place during rapid-responses of cells to stimuli, as RNA-seq only captured stable, mature RNAs. As a result, we were unable to identify or measure transcription of unstable RNAs such as enhancer RNAs (eRNAs). In 2016, Mahat *et al.* published a technique called PRO-seq to identify and quantify nascent RNAs produced globally by determining the exact location of Pol II with a single base-pair resolution during the response, without using any antibody. This bridged the gap between identification of stable RNAs versus unstable RNAs and accounted for the massive changes in transcription taking place in intergenic regions.

However, changes in RNA is not the only output of rapid response. To fully understand the scope of rapid response to a stimulus, one needs to measure the changes in transcription factor binding, changes in various histone marks associated with transcription, changes in chromatin openness, and changes in the 3D genome organization along with changes in transcription. In our study we followed a combinatorial approach and have used PRO-seq, RNA-seq, ChIP-seq, and ATAC-seq to measure rapid changes in transcriptome in human cells exposed to various stressors.

In our study with mammalian breast cancer MCF7 and mammalian leukemic K562 cells treated with heat-dependent (heat-shock at 43°C) and heat-independent stimuli (Arsenic) inducing the same HSR pathway, we show that genome-wide binding of the classical heat-shock response transcription factor HSF1 is more sensitive to these stimuli than changes in genome-wide transcription. Overall, gene transcription changes by ~ 30% whereas number of HSF1 binding sites increases by ~88%. This surprising observation indicates that instead of focusing on just transcriptional changes, one should also look at transcription factors as a measurable output of cell's response to a stimulus. Whether this increase in HSF1 binding leads to regulation of intergenic transcription or is necessary for conformational changes required for transcription are questions that need to be addressed in future.

The similarity in numbers of HSF1-bound regions versus transcriptional changes in both cell types and conditions made us wonder if universal mechanisms are at play that do not involve 'micromanaging' individual loci and is true for all HSF1-mediated HSR. Previous studies by Susan Lindquist and colleagues (Mendillo *et al*, 2012)[137] suggested of an HS-independent program that is present in highly malignant cancer cells. Cancer cells

with high metastatic potential ectopically activate HSR pathways to activate stress-responsive genes in the absence of stress as an important adaptation mechanism for reprogramming their metabolism. Mendillo *et al,* proposed of an additional program which is stress-dependent and leads to activation of short-term stress response in cells. This and other studies revealed an unexpectedly high complexity of HSR program.

Addressing the question of whether the response of cells to a stimulus is determined conditionally by the stimulus or is rigidly specified by pre-existing interactions at the ground state of the cells was critical. Our studies performed on two distinct cell types with two distinct HSR-inducing stimuli clearly show that low-grade tumor cells follow the same HSR program in heat-dependent and heat-independent stress. HSR mediated by HS or As have more than 50% HSF1 binding sites in common and is true for both cell types.

Having established that there is one common mechanism that take place within a cell irrespective of heat-dependent or heat-independent stimuli, we next asked if the mechanism of response is similar between the two types of cells. To our surprise, HSR between the two cell types are quite distinct. Although some of the genes activated in the two cell types in response to HSR are different, a common core of HS-responsive genes is activated irrespective of the stresses or the cell types. However, the major difference lies not in the differentially expressed genes but in the regions of HSF1 enrichment. The two cells show distinct HSF1 binding patterns to the same stress not at gene promoters, but at the distal intergenic regions.

We further found HSF1 binding is specific and driven by its motif in both cell-types and both treatments. This result shows that HSF1 binding is flexible and its binding sites depends on the cell type irrespective of the stimulus. In both cell-types, HSF1 binds only

to ~1% of its potential binding sites. This led us to ask the next question of how HSF1 selects its binding sites. We hypothesized HSF1 binds to specific motifs at open chromatin regions, as HSF1 has not been shown to properties of pioneer factors. However, our hypothesis was proved false and we found no correlation between HSF1 binding and chromatin accessibility.

There might be quite a few possibilities of how HSF1 finds its binding site. Deriving from our results and existing literature, two of the most probable mechanisms are outlined below:

a) HSF1 might be acting with a co-transcription factor in mammalian cells that is located in close proximity to HSF1 binding site and aids HSF1 to select specific binding sites either through motif recognition or by loosening of DNA or by both. Further experiments need to be conducted to look at stress-responsive trans-activating domains (TADs) which might serve as potential binding sites for co-transcription factors and their interaction with HSF1-binding sites during HSR. Recently it has been shown that HSR in K562 cells do not lead to TAD re-arrangements, meaning that these chromatin boundaries might have been already established at ground state of cells and do not get rapidly reprogrammed every time a stimulus hits a cell (Ray *et al*., 2019)[138].

b) Another possibility of how HSF1 selects its binding sites might be similar to what has been proposed by Chen *et al* [139] in the context of transcription factor Sox2. Cells might harbor two populations of HSF1 - one that is stably bound to chromatin at ground state and the second one is rapidly induced with stress. This second HSF1 population follows the trail of the already-bound HSF1 at ground state and binds to

those specific sites based on cognate-sequence and the hierarchical order set by the stably associated HSF1. With the available state-of-the-art techniques such as ChIP-seq for determination of genome-wide DNA-binding molecules, it is difficult to detect low-intensity chromatin-bound HSF1. However, if we compare HSF1-bound regions in stressed conditions versus unstressed conditions, we detect most (> 80%) of peaks that were present at NHS are still present at HS or As.

Considering HSF1 binding and genome-wide changes in transcription as two hallmarks of HSR, we wondered if mechanisms of Pol II regulation follow a universal pattern or if they are condition- or cell-type specific. Pol II ChIP-seq analysis on MCF7 cells in unstressed vs stressed conditions reveal that more than 80% of Pol II bound regions at ground state remain unchanged with HSR. This is consistent with the concept of establishment of transcriptional hubs during lineage determination. Previous studies by Mitchell and Fraser, 2007 have found that cells have Pol II factories present at uninduced conditions which are ready to respond to any kind of stimuli. Unlike what has been reported by Heida *et al,* 2005 in Chinese hamster ovary cells, we do not observe complete dissociation of chromatin-bound Pol II and re-establishment of Pol II factories upon heat-shock. However, a slight decrease in chromatin-bound Pol II at the intergenic regions with HS or As treatment has been noted in MCF7 cells.

One of the hallmarks of HSR is massive repression of transcription. Repression upon heat shock has been noted in *Drosophila* [98] and in recent studies of mouse [103]and human [113]. When we considered HSR in MCF7 cells, we observed a similar number of genes that are activated with heat shock as is reported. However, we did not observe a massive repression and transcription remained unabated in MCF7 cells.

Transcriptional activation of gene expression occurs by increased recruitment and release of paused Pol II into gene body. This mechanism of activation is conserved in all cell types under all conditions. However, mechanisms of repression appear to be cell-type specific. K562 repress gene expression by blocking release of Pol II into gene body without any effect on the recruitment of Pol II at promoters. MCF7 cells, on the other hand, repress gene transcription by decreased recruitment of Pol II at promoters. Previous studies on MCF7 cells induced with β-estrogen (E_2) also show repression of genes by blocking Pol II initiation. Thus, our study and previous studies on different mammalian cells including MCF7 and K562 strongly suggest different cell lines induce transcriptional changes by acting on distinct rate-limiting steps – some affecting recruitment versus other affect release.

The answer to the question why MCF7 cells exhibit decreased Pol II recruitment versus K562 has no effect on recruitment but blocks pause release might be in the availability of factors that control Pol II recruitment, pause initiation and pause release. Studies have shown that lack of active pause release factors such as P-TEFb and Mediator complex during HSR can lead to accumulation of paused Pol II at promoters. Another important factor of pause-regulation, NELF which initiates pausing have not been much looked at in the context of HSR. Our future experiments investigate if one of the pause-mediating factors such as NELF is knocked down, would that increase repression in MCF7 cells under heat-shock and provide a justification for the difference in transcriptional mechanism in MCF7 cells.

When Pol II dynamics was observed at earlier time-points such as 12 min and 30 min, a similar mechanism of paused Pol II retention was observed, as is seen in K562. At

30 min of HS, MCF7 cells show a higher repression of genes as well. These results indicate that MCF7 cells might adapt to HSR faster than K562. We propose that initially MCF7 cells start repressing gene expression by increasing Pol II pausing. Once it reaches a certain threshold of accumulated Pol II at promoter or lack of free usable Pol II in the nucleus, promoter-bound Pol II is released either into the gene-body or escapes from chromatin into the unbound Pol II pool. This Pol II is then used up either by upregulated genes or for genes to resume basal level of transcription leading to a decrease in repression. Whether this response of MCF7 makes the cells more susceptible to accumulate mutations in proteins expressed during stress response is a question that needs to be tested.

As mentioned previously, stress responses not only affect transcription regulatory proteins and gene expression, but also affects regulatory regions such as enhancers. In a different system using inflammatory cytokine IL-1β as a stress inducer in human lung cancer cells A549, we measured transcriptional changes in the regulatory regions of DNA.

Comparison of promoter-proximal versus gene-body transcription reveal major disconnect between transcription initiation and transcription elongation. ~ 72% of genes that start transcription in response to IL-1β end up with abortive transcripts and might be undergoing premature termination. Such high rate of abortive transcription indicates that all genes with transcription initiation machinery probably start rapid transcription with IL-1β stimulus. Specificity of the response is determined later during pausing whether the gene will be activated or remain non-responsive.

This observation raises the question how genes (both promoter-proximal and gene-body) are prioritized for activation or repression. We hypothesized that presence of an active enhancer region in close proximity might serve as a template for determining pattern

of gene expression. In fact, we found 62% of promoter-proximal regions and 57% of gene-body regions showed the same pattern of expression as their associated enhancers.

This result made us wonder about the timing of transcription and if eRNA is transcribed before mRNA setting up the pattern for mRNA expression. However, short-time course PRO-seq studies and RT qPCR against KLF6 eRNA and mRNA refute the idea and reveal that these RNAs are co-transcribed without any delay.

In essence, stress response circuits are most likely organized during cell lineage determination and stress-response elements belonging to common signaling pathways follow similar mechanisms of action. Dynamic changes in cells to stress are most pronounced in the intergenic regions or at the potential enhancers that show differential eRNA transcription and serve as transcription factor binding sites. Rapid response to stress does not lead to major genomic reorganization or changes in histone proteins and remains stable during stress response.

REFERENCES

1. Hahn, S. Structure and mechanism of the RNA Polymerase II transcription machinery. Nat Struct Mol Biol **11**, (2004).

2. Hampsey, M. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* **62**, 465–503 (1998).

3. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166**, 358–368 (2016).

4. Levens, D. & Larson, D. R. A new twist on transcriptional bursting. *Cell* **158**, 241–242 (2014).

5. Larsson, A. J. M. *et al.* Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251–254 (2019).

6. Bartman, C. R. *et al.* Transcriptional Burst Initiation and Polymerase Pause Release Are Key Control Points of Transcriptional Regulation. *Mol. Cell* **73**, 519-532.e4 (2019).

7. Chong, S., Chen, C., Ge, H. & Xie, X. S. Mechanism of transcriptional bursting in bacteria. *Cell* **158**, 314–326 (2014).

8. Chen, H. & Larson, D. R. What have single-molecule studies taught us about gene expression? *Genes Dev.* **30**, 1796–810 (2016).

9. Blake, W. J. *et al.* Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24**, 853–65 (2006). 10. Hornung, G. *et al.* Noise-mean relationship in mutated promoters. *Genome Res.* **22**, 2409–17 (2012).

11. Brouwer, I. & Lenstra, T. L. Visualizing transcription: key to understanding gene expression dynamics. *Curr. Opin. Chem. Biol.* **51**, 122–129 (2019).

12. Fraser, N. W., Sehgal, P. B. & Darnell, J. E. DRB-induced premature termination of late adenovirus transcription. *Nature* **272**, 590–593 (1978).

13. Gariglio, P., Bellard, M. & Chambon, P. Clustering of RNA polymerase B molecules in the 5' moiety of the adult beta-globin gene of hen erythrocytes. *Nucleic Acids Res.* **9**, 2589–98 (1981).

14. Gilmour, D. S. & Lis, J. T. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. *Mol. Cell. Biol.* **6**, 3984–9 (1986).

15. Rougvie, A. E. & Lis, J. T. Postinitiation transcriptional control in Drosophila melanogaster. *Mol. Cell. Biol.* **10**, 6041–5 (1990).

16. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–31 (2012).

17. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–8 (2008).

18. Min, I. M. *et al.* Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev.* **25**, 742–754 (2011).

19. Larschan, E. *et al.* X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. *Nature* **471**, 115–118 (2011).

20. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 167–77 (2015).

21. Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **3**, (2014).

22. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science (80-.).* **339**, 950–953 (2013).

23. Scheidegger, A. *et al.* Genome-wide RNA pol II initiation and pausing in neural progenitors of the rat. *BMC Genomics* **20**, 477 (2019).

24. Liu, X., Kraus, W. L. & Bai, X. Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends Biochem. Sci.* **40**, 516–25 (2015).

25. Gilchrist, D. A. *et al.* NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev.* **22**, 1921–33 (2008).

26. Price, D. H. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol. Cell. Biol.* **20**, 2629–34 (2000).

27. Luo, Z., Lin, C. & Shilatifard, A. The super elongation complex (SEC) family in transcriptional control. *Nat. Rev. Mol. Cell Biol.* **13**, 543–547 (2012).

28. Narita, M. *et al.* Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell* **113**, 703–16 (2003).

29. Chen, F. X. *et al.* PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. *Cell* **162**, 1003–15 (2015).

30. Samarakkody, A. *et al.* RNA polymerase II pausing can be retained or acquired during activation of genes involved in the epithelial to mesenchymal transition. *Nucleic Acids Res.* **43**, 3938–49 (2015).

31. Gilchrist, D. A. *et al.* Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* **143**, 540–51 (2010).

32. Veloso, A. *et al.* Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* **24**, 896–905 (2014).

33. Danko, C. G. *et al.* Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol. Cell* **50**, 212–22 (2013).

34. Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D. L. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat. Struct. Mol. Biol.* **15**, 71–78 (2008).

35. Saponaro, M. *et al.* RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell* **157**, 1037–49 (2014).

36. Shilatifard, A., Conaway, R. C. & Conaway, J. W. The RNA Polymerase II Elongation Complex. *Annu. Rev. Biochem.* **72**, 693–715 (2003).

37. Greger, I. H. & Proudfoot, N. J. Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of Saccharomyces cerevisiae. *EMBO J.* **17**, 4771–4779 (1998).

38. SHEARWIN, K., CALLEN, B. & EGAN, J. Transcriptional interference – a crash course. *Trends Genet.* **21**, 339–345 (2005).

39. Proudfoot, N. J. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science (80-.).* **352**, aad9926–aad9926 (2016).

40. Gullerova, M. & Proudfoot, N. J. Cohesin Complex Promotes Transcriptional Termination between Convergent Genes in S. pombe. *Cell* **132**, 983–995 (2008).

41. Prescott, E. M. & Proudfoot, N. J. Transcriptional collision between convergent genes in budding yeast. *Proc. Natl. Acad. Sci.* **99**, 8796–8801 (2002).

42. Hobson, D. J., Wei, W., Steinmetz, L. M. & Svejstrup, J. Q. RNA Polymerase II Collision Interrupts Convergent Transcription. *Mol. Cell* **48**, 365–374 (2012).

43. Porrua, O. & Libri, D. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* **16**, 190–202 (2015).

44. Sheridan, R. M., Fong, N., D'Alessandro, A. & Bentley, D. L. Widespread Backtracking by RNA Pol II Is a Major Effector of Gene Activation, 5′ Pause Release, Termination, and Transcription Elongation Rate. *Mol. Cell* **73**, 107-118.e4 (2019).

45. Chapman, R. D., Conrad, M. & Eick, D. Role of the mammalian RNA polymerase II C-terminal domain (CTD) nonconsensus repeats in CTD stability and cell proliferation. *Mol. Cell. Biol.* **25**, 7665–74 (2005).

46. Jiang, Y., Yan, M. & Gralla, J. D. A three-step pathway of transcription initiation leading to promoter clearance at an activation RNA polymerase II promoter. *Mol. Cell. Biol.* **16**, 1614–21 (1996).

47. Mayfield, J. E., Burkholder, N. T. & Zhang, Y. J. Dephosphorylating eukaryotic RNA polymerase II. *Biochim. Biophys. Acta* **1864**, 372–87 (2016).

48. Takaku, M. *et al.* GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. *Genome Biol.* **17**, 36 (2016).

49. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10**, 252–263 (2009).

50. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–51 (2013).

51. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–26 (2012).

52. Sandmann, T. *et al.* A core transcriptional network for early mesoderm development in Drosophila melanogaster. *Genes Dev.* **21**, 436–49 (2007).

53. Lebrecht, D. *et al.* Bicoid cooperative DNA binding is critical for embryonic patterning in Drosophila. *Proc. Natl. Acad. Sci.* **102**, 13176–13181 (2005).

54. Miller, J. A. & Widom, J. Collaborative competition mechanism for gene activation in vivo. *Mol. Cell. Biol.* **23**, 1623–32 (2003).

55. Falvo, J. V, Thanos, D. & Maniatis, T. Reversal of intrinsic DNA bends in the IFN beta gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell* **83**, 1101–11 (1995).

56. Panne, D., Maniatis, T. & Harrison, S. C. An Atomic Model of the Interferon-β Enhanceosome. *Cell* **129**, 1111–1123 (2007).

57. Cirillo, L. A. & Zaret, K. S. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol. Cell* **4**, 961–9 (1999).

58. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).

59. Yie, J., Senger, K. & Thanos, D. Mechanism by which the IFN-beta enhanceosome activates transcription. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13108–13 (1999).

60. Merika, M., Williams, A. J., Chen, G., Collins, T. & Thanos, D. Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. *Mol. Cell* **1**, 277–87 (1998).

61. Thanos, D. & Maniatis, T. Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome. *Cell* **83**, 1091–1100 (1995).

62. Cohen, I., Poręba, E., Kamieniarz, K. & Schneider, R. Histone modifiers in cancer: friends or foes? *Genes Cancer* **2**, 631–47 (2011).

63. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–80 (2001).

64. Batie, M. *et al.* Hypoxia induces rapid changes to histone methylation and reprograms chromatin. *Science* **363**, 1222–1226 (2019).

65. Lawrence, M., Daujat, S. & Schneider, R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet.* **32**, 42–56 (2016).

66. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* **107**, 21931–21936 (2010).

67. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**, 621–637 (2018).

68. Shandilya, J. & Roberts, S. G. E. The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819**, 391–400 (2012).

69. Deng, W. & Roberts, S. G. E. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev.* **19**, 2418–2423 (2005).

70. Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D. & Ebright, R. H. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* **12**, 34–44 (1998).

71. Ponjavic, J. *et al.* Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol.* **7**, R78 (2006).

72. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–86 (2014).

73. Cho, K. W. Y. Enhancers. *Wiley Interdiscip. Rev. Dev. Biol.* **1**, 469–78 (2012).

74. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).

75. Farley, E. K. *et al.* Suboptimization of developmental enhancers.

76. Farley, E. K., Olson, K. M., Zhang, W., Rokhsar, D. S. & Levine, M. S. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6508–13 (2016).

77. Farley, E. K. *et al.* Suboptimization of developmental enhancers. *Science* **350**, 325–8 (2015).

78. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825–37 (2013).

79. Yang, Y. *et al.* Enhancer RNA-driven looping enhances the transcription of the long noncoding RNA DHRS4-AS1, a controller of the DHRS4 gene cluster. *Sci. Rep.* **6**, 20961 (2016).

80. Enhancer-derived RNA: A Primer. *Genomics. Proteomics Bioinformatics* **15**, 196–200 (2017).

81. Li, W., Notani, D. & Rosenfeld, M. G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.* **17**, 207–223 (2016).

82. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–20 (2013).

83. Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–7 (2010).

84. Hah, N. *et al.* Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E297-302 (2015).

85. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–51 (2013).

86. Kron, K. J., Bailey, S. D. & Lupien, M. Enhancer alterations in cancer: a source for a cell identity crisis. *Genome Med.* **6**, 77 (2014).

87. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2015).

88. Brown, J. D. *et al.* NF-κB Directs Dynamic Super Enhancer Formation in Inflammation and Atherogenesis. *Mol. Cell* **56**, 219–231 (2014).

89. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. (2013). doi:10.1016/j.cell.2013.09.053

90. Zhang, X. *et al.* Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat. Genet.* **48**, 176–82 (2016).

91. Barolo, S. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays* **34**, 135–41 (2012).

92. Cannavò, E. *et al.* Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr. Biol.* **26**, 38–51 (2016).

93. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–12 (2009).

94. Joo, J.-Y., Schaukowitch, K., Farbiak, L., Kilaru, G. & Kim, T.-K. Stimulus-specific combinatorial functionality of neuronal c-fos enhancers. doi:10.1038/nn.4170

95. Cinghu, S. *et al.* Intragenic Enhancers Attenuate Host Gene Expression. *Mol. Cell* **68**, 104-117.e6 (2017).

96. Stetter, K. O. History of discovery of the first hyperthermophiles. *Extremophiles* **10**, 357–362 (2006).

97. Richter, K., Haslbeck, M. & Buchner, J. Molecular Cell Review The Heat Shock Response: Life on the Verge of Death. *Mol. Cell* **40**, 253–266 (2010).

98. Lindquist, S. *THE HEAT-SHOCK RESPONSE*. (1986).

99. Jamrich, M., Greenleaf, A. L. & Bautz, E. K. F. *Localization of RNA polymerase in polytene chromosomes of Drosophila melanogaster (RNA polymerase B and histone HI antibodies/indirect immunofluorescence/bands and interbands)*. *Genetics* **74**, (1977).

100. Ritossa, F. A new puffing pattern induced by temperature shock and DNP in drosophila. *Experientia* **18**, 571–573 (1962).

101. Åkerfelt, M., Morimoto, R. I. & Sistonen, L. Heat shock factors: integrators of cell stress, development and lifespan. *Nat Rev Mol Cell Biol* **11**, 545–555 (2010).

102. Kantidze, O. L., Velichko, A. K. & Razin, S. V. MINIIREVIEW Heat StresssInduced Transcriptional Repression. **80**, 118111185 (2015).

103. Mahat, D. B., Salamanca, H. H., Duarte, F. M., Danko, C. G. & Lis, J. T. Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation Correspondence Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. (2016). doi:10.1016/j.molcel.2016.02.025

104. Zhong, M., Orosz, A. & Wu, C. Direct sensing of heat and oxidation by Drosophila heat shock transcription factor. *Mol. Cell* **2**, 101–8 (1998).

105. Larson, J. S., Schuetz, T. J. & Kingston, R. E. In vitro activation of purified human heat shock factor by heat. *Biochemistry* **34**, 1902–11 (1995).

106. Goodson, M. L. & Sarge, K. D. Heat-inducible DNA binding of purified heat shock transcription factor 1. *J. Biol. Chem.* **270**, 2447–50 (1995).

107. Steurer, C. *et al.* HSF1 mediated stress response of heavy metals. *PLoS One* **13**, e0209077 (2018).

108. Dukler, N. *et al.* Nascent RNA sequencing reveals a dynamic global transcriptional response at genes and enhancers to the natural medicinal compound celastrol. (2017). doi:10.1101/gr.222935.117

109. Workman, J. L. Nucleosome displacement in transcription. *Genes Dev.* **20**, 2009–17 (2006).

110. Schwartz, B. E. & Ahmad, K. Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes Dev.* **19**, 804–14 (2005).

111. Zhao, J., Herrera-Diaz, J. & Gross, D. S. Domain-wide displacement of histones by activated heat shock factor occurs independently of Swi/Snf and is not correlated with RNA polymerase II density. *Mol. Cell. Biol.* **25**, 8985–99 (2005).

112. Hieda, M., Winstanley, H., Maini, P., Iborra, F. J. & Cook, P. R. *Different populations of RNA polymerase II in living mammalian cells*.

113. Vihervaara, A. *et al.* Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nat. Commun.* **8**, 255 (2017).

114. Lee, G., Walser, T. C. & Dubinett, S. M. Chronic inflammation, chronic obstructive pulmonary disease, and lung cancer. *Curr. Opin. Pulm. Med.* **15**, 303–7 (2009).

115. O'callaghan, D. S., Dearbhaile O'donnell, †, O'connell, F. & O'byrne, K. J. *The Role of Inflammation in the Pathogenesis of Non-small Cell Lung Cancer*. (2010). doi:10.1097/JTO.0b013e3181f387e4

116. Engels, E. A. Inflammation in the development of lung cancer: epidemiological evidence. *Expert Rev. Anticancer Ther.* **8**, 605–615 (2008).

117. Parkin, D. M. The global health burden of infection-associated cancers in the year 2002. *Int. J. Cancer* **118**, 3030–3044 (2006).

118. Bremnes, R. M. *et al.* The Role of Tumor-Infiltrating Immune Cells and Chronic Inflammation at the Tumor Site on Cancer Development, Progression, and Prognosis: Emphasis on Non-small Cell Lung Cancer. *J. Thorac. Oncol.* **6**, 824–833 (2011).

119. Domingues, P. *et al.* Tumor infiltrating immune cells in gliomas and meningiomas. *Brain. Behav. Immun.* **53**, 1–15 (2016).

120. Zheng, X., Hu, Y. & Yao, C. The paradoxical role of tumor-infiltrating immune cells in lung cancer. *Intractable rare Dis. Res.* **6**, 234–241 (2017).

121. Lewis, A. M., Varghese, S., Xu, H. & Alexander, H. R. Interleukin-1 and cancer progression: the emerging role of interleukin-1 receptor antagonist as a novel therapeutic agent in cancer treatment. *J. Transl. Med.* **4**, 48 (2006).

122. Van Eeden, S., Leipsic, J., Paul Man, S. F. & Sin, D. D. The Relationship between Lung Inflammation and Cardiovascular Disease. *Am. J. Respir. Crit. Care Med.* **186**, 11–16 (2012).

123. Engström, G. *et al.* Lung Function and Cardiovascular Risk Relationship With Inflammation-Sensitive Plasma Proteins. doi:10.1161/01.CIR.0000037220.00065.0D

124. Suwa, T. *et al.* Particulate air pollution induces progression of atherosclerosis. *J. Am. Coll. Cardiol.* **39**, 935–42 (2002).

125. Smeeth, L. *et al.* Risk of Myocardial Infarction and Stroke after Acute Infection or Vaccination. *N. Engl. J. Med.* **351**, 2611–2618 (2004).

126. Corrales-Medina, V. F. *et al.* Cardiac Complications in Patients With Community-Acquired Pneumonia: Incidence, Timing, Risk Factors, and Association With Short-Term Mortality. *Circulation* **125**, 773–781 (2012).

127. Phrommintikul, A. *et al.* Influenza vaccination reduces cardiovascular events in patients with acute coronary syndrome.

128. Nichol, K. L. *et al.* Influenza Vaccination and Reduction in Hospitalizations for Cardiac Disease and Stroke among the Elderly. *N. Engl. J. Med.* **348**, 1322–1332 (2003).

129. Kullo, I. J., Edwards, W. D. & Schwartz, R. S. Vulnerable plaque: pathobiology and clinical implications. *Ann. Intern. Med.* **129**, 1050–60 (1998).

130. Coulter, K. R., Wewers, M. D., Lowe, M. P. & Knoell, D. L. Extracellular Regulation of Interleukin (IL)-1 β through Lung Epithelial Cells and Defective IL-1 Type II Receptor Expression. *Am. J. Respir. Cell Mol. Biol.* **20**, 964–975 (1999).

131. Weber, A., Wasiliew, P. & Kracht, M. Interleukin-1 (IL-1) Pathway. *Sci. Signal.* **3**, cm1–cm1 (2010).

132. Dinarello, C. A. *Biologic Basis for Interleukin-l in Disease*. *The Journal of The American Society of Hematology* **87**,

133. Arondel, J., Singer, M., Matsukawa, A., Zychlinsky, A. & Sansonetti, P. J. Increased interleukin-1 (IL-1) and imbalance between IL-1 and IL-1 receptor antagonist during acute inflammation in experimental Shigellosis. *Infect. Immun.* **67**, 6056–66 (1999).

134. Lopalco, G. *et al.* Interleukin-1 as a Common Denominator from Autoinflammatory to Autoimmune Disorders: Premises, Perils, and Perspectives. *Mediators Inflamm.* **2015**, 1–21 (2015).

135. Tullai, J. W. *et al.* Immediate-early and delayed primary response genes are distinct in function and genomic architecture. *J. Biol. Chem.* **282**, 23981–95 (2007).

136. Syafruddin, S. E. *et al.* A KLF6-driven transcriptional network links lipid homeostasis and tumour growth in renal carcinoma. doi:10.1038/s41467-019-09116-x

137. Mendillo, M. L. *et al.* HSF1 Drives a Transcriptional Program Distinct from Heat Shock to Support Highly Malignant Human Cancers. (2012). doi:10.1016/j.cell.2012.06.031

138. Ray, J. *et al.* Chromatin conformation remains stable upon extensive transcriptional changes driven by heat shock. *Proc. Natl. Acad. Sci.* **116**, 19431–19439 (2019).

139. Chen, J. *et al.* Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells. (2014). doi:10.1016/j.cell.2014.01.062